DOCTOR OF PHILOSOPHY

Global identification of SUMO modification sites

Tammsalu, Triin

Award date:
2016

Global identification of SUMO modification sites

Triin Tammsalu

October 2016

A thesis submitted to the University of Dundee in partial fulfilment of the requirements for the degree of Doctor of Philosophy

Declaration

I declare that I am the author of this thesis and that, unless otherwise stated, all references cited have been consulted; that the work of which this thesis is a record of has been performed by me, and that it has not been previously accepted for a higher degree: where the thesis is based upon joint research, the nature and extent of my individual contribution is defined.

Triin Tammsalu

October 2016

I certify that Triin Tammsalu performed the work for which this thesis is a record. The conditions of the relevant Ordinance and Regulations have been fulfilled.

Professor Ronald T. Hay

Acknowledgements

This thesis would not exist without the support and guidance of:

Prof. Ronald T. Hay

Mike, Ivan, Fede, Ellis, Adel

RTH group

Mom, Dad

Richard

Jaanus, Liisi, Marita, Margit, Toomas

Südamekommuun

Eva, Adriana, Magda

MMB people

Christian Cole

Genevieve Thon, Julie B. Køhler

Prof. Angus I. Lamond, Andrea Pawellek

Prof. Carol MacKintosh, Eric Griffis, Prof. Frank Sargent

EU 7th Framework Programme ITN UPStream

Wellcome Trust

Abstract

Posttranslational modification by Small Ubiquitin-like modifiers (SUMOs) affects the behaviour of proteins involved in diverse cellular processes and is indispensable for cell survival. Proteomic studies have identified hundreds of putative targets of SUMO conjugation, but understanding the substrate-specific consequences of the modification requires knowledge of precise target lysines. Procedures for the high-throughput mapping of SUMO modification sites, however, are lacking.

A method has been developed enabling global identification of sumoylated lysines by mass spectrometry (MS). The workflow entails *in vivo* conjugation to hexahistidine (6His)-tagged SUMO, in which the residue preceding the C-terminal Gly-Gly (diGly) has been mutated to lysine (6His-SUMO^{KGG}). Lys-C digestion of 6His-SUMO^{KGG} conjugates yields a diGly remnant on modified lysines that cannot be attributed to any other known modifier. This allows enrichment of SUMO remnant-modified peptides with a diGly-Lys (K-ε-GG)-specific antibody prior to their MS-based analysis.

Application of the workflow enabled the unambiguous identification of 612 SUMO2 conjugation sites from the nuclear fraction of unperturbed human cells and 8262 sites from the same fraction of heat-stressed cells. Exposure to proteotoxic stress increased the extent of substrate sumoylation, and targeted proteins with similar functions or found within the same macromolecular complexes. These data provide a comprehensive resource for future research on the physiological role of sumoylation in health and disease.

Table of Contents

Table of Figures

Table of Tables

Table of Equations

Amino acids

Name	Three-letter abbreviation	One-letter abbreviation
Alanine	Ala	A
Arginine	Arg	R
Asparagine	Asn	N
Aspartic acid	Asp	D
Cysteine	Cys	C
Glutamine	Gln	Q
Glutamic acid	Glu	E
Glycine	Gly	G
Histidine	His	H
Isoleucine	Ile	I
Leucine	Leu	L
Lysine	Lys	K
Methionine	Met	M
Phenylalanine	Phe	F
Proline	Pro	P
Serine	Ser	S
Threonine	Thr	T
Tryptophan	Trp	W
Tyrosine	Tyr	Y
Valine	Val	V
Hydrophobic	—	ψ
Any	—	x

Abbreviations

6His	Hexahistidine
10His	Decahistidine
AMP	Adenosine monophosphate
APP	Amyloid precursor protein
Appl.	Application
ATP	Adenosine triphosphate
Aβ	Amyloid-β
BCA	Bicinchoninic acid
BS3	Bis(sulfosuccinimidyl)suberate
BSA	Bovine serum albumin
C-	Carboxyl-
Cat.	Catalog
COP9	Constitutive photomorphogenesis 9
DAPI	4',6-diamidino-2-phenylindole
DeSI	Desumoylating isopeptidase
diGly	Gly-Gly
DMEM	Dulbecco's modified Eagle medium
DMP	Dimethyl pimelimidate
DMSO	Dimethyl sulfoxide
DPBS	Dulbecco's phosphate buffered saline
DTT	Dithiothreitol
E2-25K	Ubiquitin-conjugating enzyme E2 K
ECL	Enhanced chemiluminescence
eYFP	Enhanced Yellow fluorescent protein
FASP	Filter aided sample preparation
FBS	Fetal bovine serum
FDR	False discovery rate
Fr.	Fraction
GTF2I	General transcription factor II-I
Gu-HCl	Guanidine hydrochloride
HA	Human influenza virus hemagglutinin
HCD	Higher energy collisional dissociation
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
hnRNP	Heterogenous nuclear ribonucleoprotein
HPA	Human Protein Atlas database
HSF1	Heat shock factor protein 1
HSP	Heat shock protein

IAP buffer	Immunoaffinity purification buffer: 50 mM NaCl, 50 mM MOPS-NaOH pH 7.2, 10 mM Na_2HPO_4
IF	Immunofluorescence microscopy
IMAC	Immobilised metal affinity chromatography
IPA	Qiagen's Ingenuity Pathway Analysis
IRF-2	Interferon regulatory factor 2
IT	Injection time
IκBα	NF-kappa-B inhibitor alpha
K-ε-GG	Gly-Gly-modified Lys
K0	L-Lys ($^{12}C_6$, $^{14}N_2$)
K8	L-Lys ($^{13}C_6$, $^{15}N_2$)
LC	Liquid chromatography
Lys-C	Lysyl endopeptidase
m/z	Mass-to-charge ratio
MCM	Minichromosome maintenance
MEF2	Myocyte-specific enhancer factor 2
MES	2-(N-morpholino)ethanesulfonic acid
mHTT	Pathogenic variant of Huntingtin
MOPS	3-(N-morpholino)propanesulfonic acid
MS	Mass spectrometry
MSMS	Tandem mass spectrometry
N-	Amino-
NF-κB	Nuclear factor NF-kappa-B
Ni^{2+}-NTA	Nickel-nitriloacetic acid
No.	Number
NONO	Non-POU domain-containing octamer-binding protein
PAGE	Polyacrylamide gel electrophoresis
PARP1	Poly [ADP-ribose] polymerase 1
PBS	Phosphate buffered saline
PcG	Polycomb group
PCNA	Proliferating cell nuclear antigen
PCR	Polymerase chain reaction
PDSM	Phosphorylation-dependent sumoylation motif
Pen-Strep	Penicillin and streptomycin
PIAS	Protein inhibitor of activated STAT
PML	Promyelocytic leukemia protein
PPAR-γ	Peroxisome proliferator-activated receptor gamma
ppm	Parts per million
PRC	Polycomb repressive complex
PRISM	Protease-reliant identification of SUMO modification

PSM	Peptide-to-spectrum match
PTM	Posttranslational modification
PVDF	Polyvinylidene difluoride
R0	L-Arg ($^{12}C_6$, $^{14}N_4$)
R10	L-Arg ($^{13}C_6$, $^{15}N_4$)
RanBP2	Ran-binding protein 2
RanGAP1	Ran GTPase-activating protein 1
RMSD	Root mean square deviation
S-	Sulfur
SAE	SUMO-activating enzyme
SAFB2	Scaffold attachment factor B2
SAX	Strong anion exchange
SDS	Sodium dodecyl sulfate
SENP	Sentrin/SUMO-specific protease
SF	Subcellular fractionation
SHMT1	Serine hydroxymethyltransferase
SILAC	Stable isotope labelling by amino acids in a cell culture
SIM	SUMO interaction motif
Smad4	Mothers against decapentaplegic homolog 4
SMEM	Minimum Essential medium Eagle
snRNP	Small nuclear ribonucleoprotein
Sp100	Nuclear autoantigen Sp-100
StageTip	Stop and Go Extraction Tip
STUbL	SUMO-targeted ubiquitin ligase
Sulfo-NHS-acetate	Sulfosuccinimidyl acetate
SUMO	Small Ubiquitin-like modifier
TAP	Tandem affinity purification
TCA	Trichloroacetic acid
TCEP	Tris(2-carboxyethyl)phosphine
TDG	G/T mismatch-specific thymine DNA glycosylase
TFA	Trifluoroacetic acid
TIC	Total ion current
TRIM28	Transcription intermediary factor 1-β
Ubl	Ubiquitin-like protein
Ufd1	Ubiquitin fusion degradation protein 1
WaLP	Wild-type α-lytic protease
WB	Western blot
YTHDF2	YTH domain-containing family protein 2

1.

Introduction

1.1. SUMO

1.1.1. Characteristics of SUMO proteins

Small Ubiquitin-like modifiers (SUMOs) are proteins named after and belonging to a family of Ubiquitin-like proteins (Ubls) (Mahajan et al., 1997). All members of the Ubl family are characterised by a remarkably similar three-dimensional core structure, which adopts a globular β-grasp ubiquitin-related fold despite the large differences in the amino acid sequences of individual proteins (Figure 1.1) (Bayer et al., 1998; Huang et al., 2004; Hochstrasser, 2009). A distinctive structural feature of SUMO is the flexible amino (N)-terminal extension that is absent in the founding member of the Ubl family: Ubiquitin (Figure 1.1A) (Bayer et al., 1998; Muller et al., 2001; Huang et al., 2004). However, most Ubls can be posttranslationally covalently bound to other proteins via their carboxyl (C)-terminus (Muller et al., 2001). At the C-terminus of most mature human Ubls are two conserved glycines known as the Gly-Gly or diGly motif (Figure 1.1A) (Flotho and Melchior, 2013).

Orthologs of SUMO proteins have been identified throughout the eukaryotic kingdoms, with most higher eukaryotes expressing multiple SUMO paralogs. In humans, three SUMO paralogs are expressed ubiquitously: SUMO1 (Small ubiquitin-related modifier 1, or Smt3C), SUMO2 (Small ubiquitin-related modifier 2, or Smt3A) and SUMO3 (Small ubiquitin-related modifier 3, or Smt3B) (Saitoh and Hinchey, 2000; Flotho and Melchior, 2013). Primary structures of mature SUMO2 and SUMO3 proteins differ by three amino acids and to date cannot be distinguished by antibodies. In

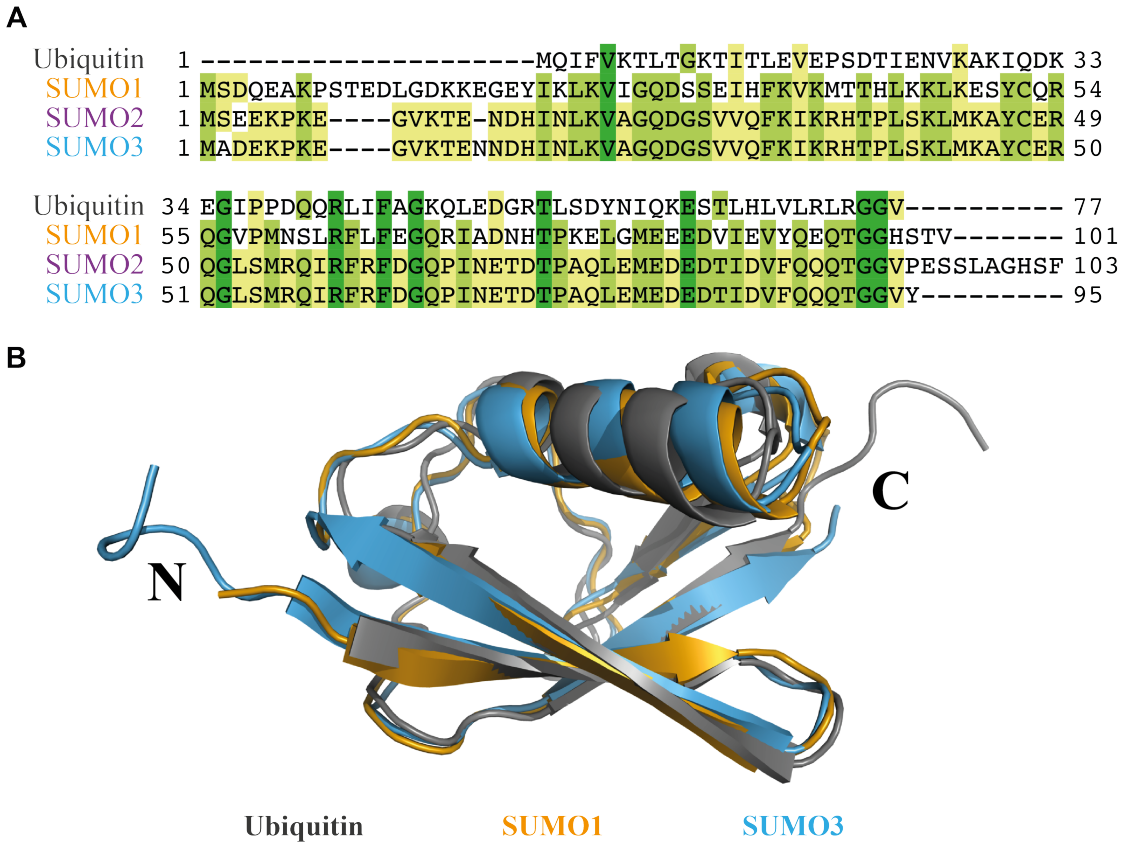

Figure 1.1 | Human Ubls share similar organisation of secondary structure elements.

(**A**) Sequence alignment of Ubiquitin, SUMO1, SUMO2 and SUMO3 coloured by identity. Amino acids conserved between two, three or all four Ubls are shown in yellow, light green or dark green, respectively. The alignment was created using Clustal Omega multiple sequence alignment tool integrated into Jalview, version 2 editor (Waterhouse et al., 2009; Sievers et al., 2011). (**B**) Structural alignment of Ubiquitin (grey; PDB entry: 1UBQ), SUMO1 (orange; PDB entry: 4WJQ), and SUMO3 (blue; PDB entry: 1WM2) (Vijay-Kumar et al., 1987; Huang et al., 2004; Cappadocia et al., 2015a). The overall root mean square deviation (RMSD) over 72 aligned residues is 1.2640 Å. The alignment was generated using a protein structure comparison service PDBeFold, version 2.59 (Krissinel and Henrick, 2004).

contrast, only ~47 % sequence identity is shared between SUMO1 and SUMO2, hence the classification of SUMO proteins into two subfamilies: SUMO1 or SUMO2/3. Western blot images of human cell lysates indicate that the majority of SUMO1 is engaged in conjugates. In comparison, a large proportion of SUMO2/3 is unconjugated and can be rapidly covalently bound to target proteins upon cell exposure to protein-

damaging stimuli, such as ethanol, hydrogen peroxide, or heat (Saitoh and Hinchey, 2000). Furthermore, SUMO has the ability to target itself for modification and polymerise into chains, although incorporation of SUMO1 into poly-SUMO2/3 chains appears to impede further chain elongation *in vitro* (Tatham et al., 2001; Matic et al., 2008; Hsiao et al., 2009).

1.1.2. SUMO conjugation pathway

Each SUMO paralog is synthesized as a precursor molecule, which is processed into an active form by cleaving the carboxyl-terminal extension and exposing the C-terminal glycine of the conserved diGly motif (Figure 1.1A). Enzymes responsible for the SUMO maturation are generically termed SUMO proteases (Hickey et al., 2012). Protein sumoylation is a process in which an isopeptide bond is formed between the C-terminal carboxylate of SUMO and the ε-amino group of a lysine residue of a substrate (Figure 1.2). The SUMO conjugation pathway is initialised by a heterodimeric E1 composed of SUMO-activating enzyme subunits 1 and 2 (SAE1/SAE2), which catalyses the adenosine triphosphate (ATP)-dependent adenylation of the C-terminal carboxyl group of SUMO. Consequent release of adenosine monophosphate (AMP) is accompanied by the formation of a thioester bond between the sulfhydryl group of the catalytic Cys residue of SAE2 and the C-terminus of SUMO (Johnson et al., 1997; Desterro et al., 1999; Gong et al., 1999; Okuma et al., 1999). Subsequently, upon interaction with E2 SUMO-conjugating enzyme Ubc9, SUMO is transesterified to the catalytic Cys residue of the E2 (Desterro et al., 1997; Johnson and Blobel, 1997). Ubc9 alone is able to directly recognise target proteins and facilitate the ultimate ligation of

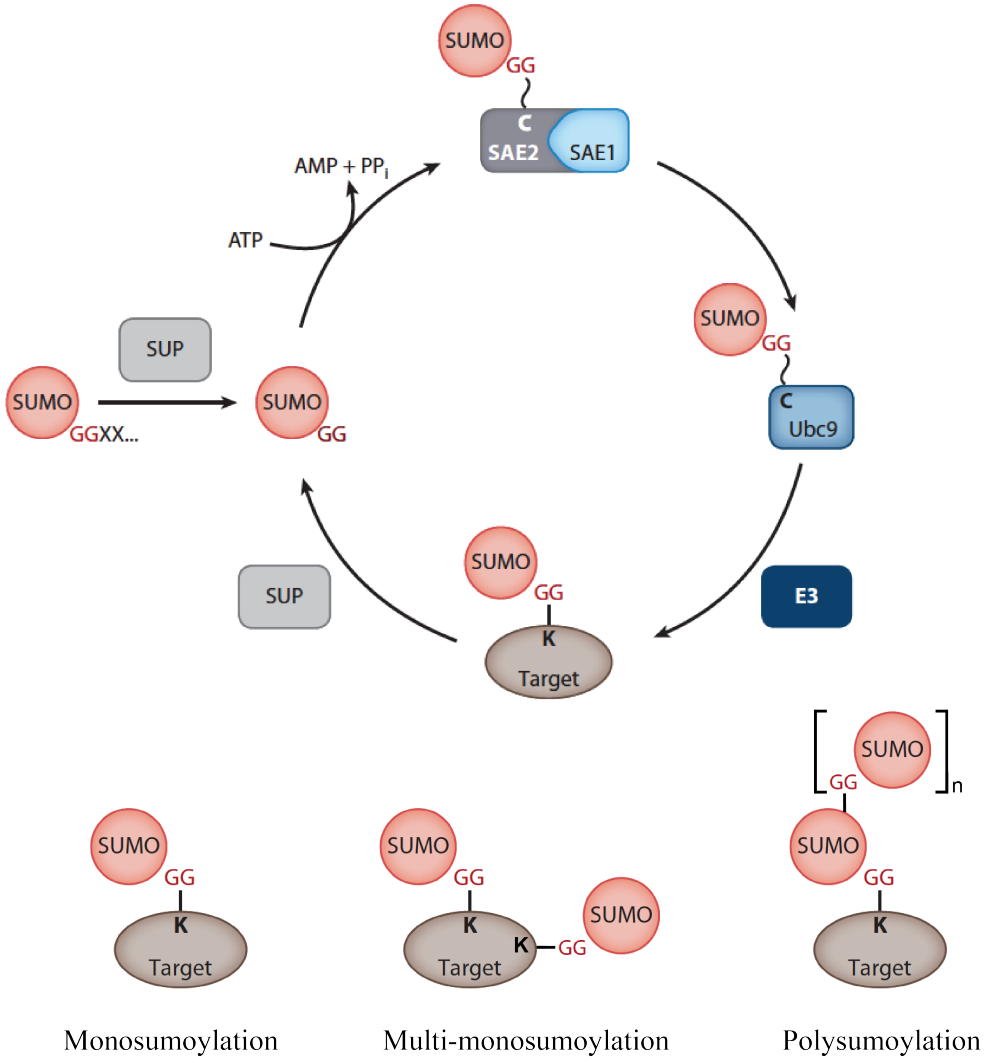

Figure 1.2 | The cycle of SUMO modification.

The nascent SUMO precursor is processed by SUMO-specific proteases (SUP) to reveal the conserved C-terminal Gly-Gly motif. The mature molecule is then activated by an E1 SUMO-activating enzyme SAE1/SAE2 and subsequently transesterified to an E2 SUMO-conjugating enzyme Ubc9. E2 facilitates the transfer of SUMO to target proteins, however the process is often catalysed by the actions of E3 SUMO-protein ligases. Substrate sumoylation can occur on a single (monosumoylation) or multiple lysine residues (multi-monosumoylation), and SUMO can form chains (polysumoylation). The cycle is completed by SUMO-specific protease-dependent desumoylation of substrates. "XX" represents the C-terminal extension of the nascent SUMO molecule; wavy and straight lines correspond to thioester or peptide bonds, respectively. Figure is adapted from (Flotho and Melchior, 2013).

the SUMO molecule to the ε-amino group of the Lys residue of the substrate (Rodriguez et al., 2001; Sampson et al., 2001; Bernier-Villamor et al., 2002). However, this step is often catalysed by E3 SUMO-protein ligases (Figure 1.2). Most E3s recruit the substrate, bind charged Ubc9 and promote SUMO transfer by coordinating the thioester in an optimal orientation for the isopeptide bond formation (Reverter and Lima, 2005; Flotho and Melchior, 2013). In a recent structural study, the molecular mechanism by which an E3 SUMO-protein ligase can neglect the specificity of Ubc9 and force-feed the target Lys into the active site of the E2, was also illustrated (Streich and Lima, 2016). Importantly, SUMO conjugation can occur on a single lysine (monosumoylation) or multiple lysine residues (multi-monosumoylation) of the same substrate protein (Figure 1.2).

The largest family of mammalian E3 SUMO-protein ligases includes NSE2 and the Protein inhibitor of activated STAT (PIAS) proteins, which are characterised by an Ubc9-binding SP-RING domain (Kahyo et al., 2001; Sachdev et al., 2001; Potts and Yu, 2005; Streich and Lima, 2016). However, a few unrelated E3s have also been identified, such as the Ran-binding protein 2 (RanBP2) (Pichler et al., 2002; Pichler et al., 2004) or the ZNF451 (Cappadocia et al., 2015b; Eisenhardt et al., 2015).

SUMO modification is reversible and proteases involved in SUMO maturation are also required for its deconjugation from substrate proteins (Hickey et al., 2012). So far, nine human SUMO-specific cysteine proteases have been identified and categorised into three structural families: six Sentrin/SUMO-specific proteases (SENP1, -2, -3, -5, -6, -7) (Yeh et al., 2000), two Desumoylating isopeptidases (DeSI-1 and -2) (Shin et al., 2012), and a SUMO-specific isopeptidase USPL1 (Schulz et al., 2012).

1.1.3. SUMO interaction motif

A requirement for SUMO to act as a molecular marker triggering the recruitment of other proteins is the ability of these downstream effectors to recognise SUMO. In general, this specificity is found in a particular sequence element known as a SUMO interaction motif (SIM). The SIM is a short repeat of branched hydrophobic residues generally conforming to one of the two sequence motifs: [V/I]x[V/I][V/I] or [V/I][V/I]x[V/I], where x represents any amino acid and residues in square brackets are interchangeable. SIMs are often N- or C-terminally flanked by a stretch of serines, acidic or both types of residues, which appear to enhance the binding affinity for SUMO. SIMs form non-covalent interactions with SUMO paralogs by adopting short β-strand conformations that insert into the groove between the β_2-strand and the α-helix of the SUMO molecule (Song et al., 2004; Song et al., 2005; Hecker et al., 2006).

Two additional binding modules, which have thus far been characterised in a limited number of proteins, form non-covalent interactions preferentially with SUMO1: the ZZ-type zinc finger domain of the E3 ubiquitin-protein ligase HERC2 and the extended arm structure of the Dipeptidyl peptidase 9 (Danielsen et al., 2012; Pilla et al., 2012).

1.1.4. Molecular mechanisms of SUMO modification

The downstream effects of protein sumoylation are often substrate specific, but in general can be categorised into three types of events. First, SUMO can compete with other Lys-targeting posttranslational modifications (PTMs), such as methylation, acetylation or conjugation to other Ubls (Figure 1.3A). For instance, transcription factor Nuclear factor NF-kappa-B (NF-κB) is activated upon the degradation of NF-kappa-B

inhibitor alpha (IκBα) via its ubiquitination at Lys[21]. However, sumoylation of Lys[21] of IκBα prevents its ubiquitination, rendering the molecule partially resistant to proteasomal degradation, and thus facilitating the inhibition of NF-κB-dependent transcription (Desterro et al., 1998). Furthermore, competition between acetylation and sumoylation has been described for Myocyte-specific enhancer factor 2A (MEF2A). In this situation, sumoylation of Lys[403] has been suggested to hinder its acetylation and lead to repression of MEF2-mediated transcription (Shalizi et al., 2006).

SUMO modification is a large PTM and can thus prevent the interactions of substrate proteins with their binding partners through steric hindrance (Figure 1.3B). For instance, sumoylation of Lys[14] of Ubiquitin-conjugating enzyme E2 K (E2-25K) competes with the binding of E2-25K to E1 Ubiquitin-like modifier-activating enzyme, and thus interferes with the thioester bond formation between the E2 and Ubiquitin (Pichler et al., 2005). Through steric hindrance, conjugation to SUMO can also inhibit the posttranslational modification of residues neighbouring the modified lysine. In the case of Cellular tumor antigen p53, sumoylation at Lys[386] appears to inhibit its subsequent *in vitro* Histone acetyltransferase p300-dependent acetylation at adjacent residues: Lys[382] and Lys[373] (Wu and Chiang, 2009).

The most prevalent molecular function of sumoylation appears to be the SUMO-SIM interaction-dependent recruitment of downstream effectors (Figure 1.3C). Although the affinities of SIMs for SUMOs are in the low micromolar range (Hecker et al., 2006), several cooperative SUMO-SIM interactions can enhance the efficiency of substrate binding. For instance, multiple SIMs are found in tandem at the N-terminal region of

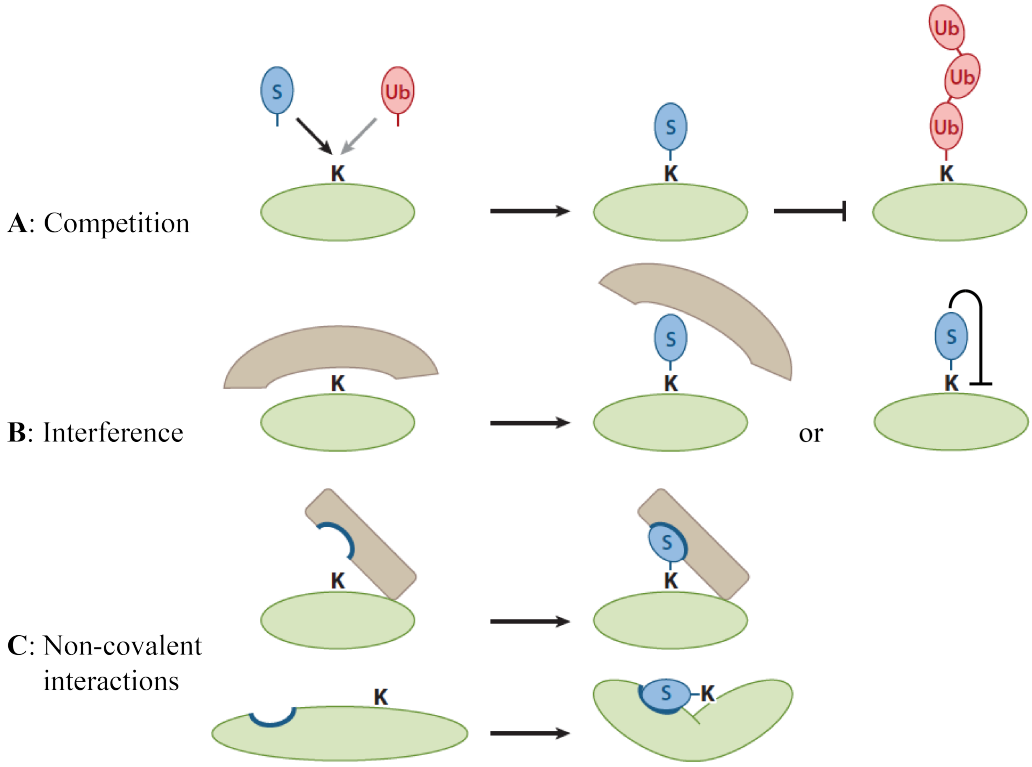

Figure 1.3 | Downstream effects of SUMO modification.

The molecular consequences of sumoylation (S; blue) are divided into three types of events. SUMO modification can: (**A**: competition) compete for acceptor lysines with other Lys-specific PTMs, such as methylation, acetylation or conjugation to Ubiquitin (Ub; red), (**B**: interference) sterically hinder either protein-protein interactions or posttranslational modification of adjacent residues, or (**C**: non-covalent interactions) bind non-covalently to inter- or intramolecular SUMO interaction motifs (SIMs) and thus induce physical associations or conformational changes. Figure is adapted from (Jentsch and Psakhye, 2013).

E3 Ubiquitin-protein ligase RNF4, which through specific binding of poly-SUMO chains promotes the ubiquitination and subsequent proteasomal degradation of several substrate proteins, including the promyelocytic leukemia protein (PML) (Tatham et al., 2008). SUMO-SIM binding can also induce intramolecular interactions. Upon sumoylation at Lys^{330}, the G/T mismatch-specific thymine DNA glycosylase (TDG) undergoes a conformational change that enables its dissociation from the abasic site after the excision of the irregular base (Steinacher and Schar, 2005).

1.2. SUMO is essential for viability

Most eukaryotic organisms require a functional sumoylation pathway for survival. The sole murine SUMO-conjugating enzyme Ubc9 is essential for the early development of mouse embryos, with Ubc9-deficient mice suffering from premature embryonic death at the post-implantation stage (Nacerddine et al., 2005). Moreover, adult mice with a conditional knock-out of Ubc9 in the gastrointestinal tract die within six days (Demarque et al., 2011). Mice deficient in SUMO1 are fertile and demonstrate no overt phenotype although increased conjugation of SUMO2/3 has been suggested to compensate for the lack of SUMO1 (Evdokimov et al., 2008; Zhang et al., 2008). Likewise, SUMO2 is dispensable for the development of mouse embryos and the appearance of postnatal mice lacking SUMO2 is indistinguishable from the wild-type littermates. In contrast, deficiency in murine SUMO3 results in severe developmental delay and has an embryonic lethal phenotype. Considering that the sequence and structure of SUMO3 is nearly identical to SUMO2, but SUMO3 is the primarily expressed paralog during mouse embryogenesis, it has been suggested that the apparent developmental differences are inflicted by an extreme decrease in the overall SUMO production (Wang et al., 2014). Inactivation of murine SENP1 leads to embryonic lethality in midgestation as a consequence of severe defects in erythro- or T and B cell lymphopoiesis (Cheng et al., 2007; Van Nguyen et al., 2012). These findings demonstrate the importance of balanced, dynamic SUMO conjugation for the viability of higher eukaryotes, and suggests at least partial functional redundancy of individual SUMO paralogs (Evdokimov et al., 2008; Wang et al., 2014).

1.3. SUMO modification in proteostasis

The homeostasis of the proteome (proteostasis) is achieved by a balanced and coordinated synthesis, folding, disaggregation, and degradation of proteins (Labbadia and Morimoto, 2015). SUMO conjugation increases under conditions of proteotoxic stress (Saitoh and Hinchey, 2000; Golebiowski et al., 2009; Tatham et al., 2011), and the modification is involved in several cellular processes required for the maintenance of proteostasis (Liebelt and Vertegaal, 2016).

Nucleotide-binding proteins are common targets of sumoylation in unperturbed and stressed human cells, and the modification of chromatin-binding proteins occurs predominantly at the promoters of actively transcribed genes (Golebiowski et al., 2009; Liu et al., 2012; Becker et al., 2013; Neyret-Kahn et al., 2013; Niskanen et al., 2015; Seifert et al., 2015). However, the general physiological role of SUMO recruitment to chromatin is still unclear, as it has been suggested to either maintain or restrain the transcription of target genes (Liu et al., 2012; Neyret-Kahn et al., 2013; Niskanen et al., 2015; Seifert et al., 2015). Coordinated sumoylation and desumoylation of substrates is also required for the biogenesis of ribosomes, where it facilitates the assembly of eukaryotic pre-ribosomal particles (Panse et al., 2006; Haindl et al., 2008; Finkbeiner et al., 2011). Furthermore, SUMO appears to be required for mRNA processing (Vassileva and Matunis, 2004; Vethantham et al., 2008; Pelisch et al., 2010; Liu et al., 2015).

Sumoylation has also been suggested to promote protein solubility, as the aggregation propensity of several natively unfolded polypeptides is reduced by the modification. Interestingly, most of these aggregation-prone polypeptides can cause protein

misfolding disorders (proteopathic diseases) (Mukherjee et al., 2009; Janer et al., 2010; Krumova et al., 2011).

The involvement of SUMO in protein degradation is mostly mediated by its crosstalk with Ubiquitin. As demonstrated in Chapter 1.1.4 competition between the Ubls for the same acceptor lysine alters the stability of IκBα. However, several other proteins have been suggested to be under similar regulation, such as the Serine hydroxymethyltransferase (SHMT1) (Anderson et al., 2012) or Mothers against decapentaplegic homolog 4 (Smad4) (Lin et al., 2003). For instance, the stability of Smad4 appears to increase through sumoylation at Lys^{113} and Lys^{159}, which prevent its Ubiquitin-mediated degradation (Lin et al., 2003).

Through SUMO-SIM interactions, polysumoylation of substrates can serve as a signal for the recruitment of SUMO-targeted E3 Ubiquitin-protein ligases (STUbLs), such as RNF4 (Tatham et al., 2008). The stability of many proteins functioning in fundamental cellular processes, including the assembly of kinetochores (Mukhopadhyay and Dasso, 2010) or survival after exposure to hyperthermia (Martin et al., 2009), hypoxia (van Hagen et al., 2010), or DNA damage (Nie and Boddy, 2016), are targets of RNF4-mediated ubiquitination and subsequent degradation. In some cases, the STUbL-dependent processing of substrates is assisted by p97 (Cdc48 in yeast)-Ufd1-Npl4 segregase, which facilitates the extraction of ubiquitinated proteins from higher order complexes (Meyer et al., 2012; Baek et al., 2013). For instance, one of the known targets of the combined action of STUbL and p97-Ufd1-Npl4 segregase is the Fanconi anemia ID complex (FANCI/FANCD2), which is removed from the sites of DNA damage by

SUMO modification, RNF4-catalysed ubiquitination, and p97-mediated extraction (Gibbs-Seymour et al., 2015).

1.4. SUMO in proteopathic diseases

The involvement of SUMO modification in human diseases is gradually becoming a topic of interest, as many proteins with a fundamental function in pathogenesis are substrates of sumoylation. Thus, SUMO conjugation has been suggested to affect carcinogenesis (Eifler and Vertegaal, 2015), cardiac disease (Mendler et al., 2016), viral infection (Everett et al., 2013), or several proteopathic diseases (Steffan et al., 2004; Eckermann, 2013; Lee et al., 2013).

1.4.1. Polyglutamine diseases

Polyglutamine diseases are a family of neurodegenerative conditions caused by an expansion of trinucleotide repeat that induces the production of a toxic aggregation-prone protein with an elongated polyglutamine track. Neurological disorders affected by the repeat expansion include spinobulbar muscular atrophy, Huntington's disease, and various types of spinocerebellar ataxias (Zoghbi and Orr, 2000). Interestingly, the solubility of several of these disease-inducing proteins is altered by conjugation to SUMO (Mukherjee et al., 2009; Janer et al., 2010).

Huntington's disease is a hereditary disorder characterised by a progressive degeneration of brain cells, resulting in involuntary movements, memory loss, and personality changes (Orr and Zoghbi, 2007). The disease is caused by an expansion of a polyglutamine repeat at the N-terminal region of Huntingtin that promotes its assembly into intermediate oligomers before inclusion into higher-order aggregates (The Huntington's Disease Collaborative Research Group, 1993; Labbadia and Morimoto, 2013). The pathogenic variant of Huntingtin (mHTT) is conjugated to

SUMO1 and SUMO3 (Steffan et al., 2004; Subramaniam et al., 2009; O'Rourke et al., 2013). Modification by SUMO3 is facilitated by the E3 SUMO-protein ligase PIAS1 and results in an accumulation of toxic insoluble mHTT (O'Rourke et al., 2013). Conjugation to SUMO1 is mediated by a GTP-binding protein Rhes and reduces its aggregation propensity (Subramaniam et al., 2009). However, in each case, the sumoylation of mHTT appears to increase cytotoxicity (Subramaniam et al., 2009; O'Rourke et al., 2013). The tissue-specific expression of Rhes could also explain the localised neuropathology of the Huntington's disease (Subramaniam et al., 2009).

1.4.2. Alzheimer's disease

Alzheimer's disease is an age-related neurodegenerative disorder manifested in patient's progressive cognitive impairment. The symptoms of the condition are inflicted by the aggregation of either a small polypeptide Amyloid-β (Aβ) or a microtubule-associated protein Tau that form plaques or neurofibrillary tangles, respectively (LaFerla et al., 2007).

Aβ is generated by the endoproteolysis of an Amyloid precursor protein (APP) via the amyloidogenic pathway (LaFerla et al., 2007). In close proximity to the endoproteolytic cleavage site of APP are two lysine residues that appear to be covalently conjugated to SUMO. Sumoylation of APP leads to the reduced accumulation of Aβ aggregates and has been suggested to sterically hinder the processing of APP (Zhang and Sarge, 2008). Remarkably, some patients with Alzheimer's disease carry a mutation that is predicted to produce an APP protein with one of the SUMO target lysines replaced by an asparagine (Mullan et al., 1992; Zhang and Sarge, 2008). However, conflicting observations have also been reported. In those studies, the overproduction of either all

SUMO paralogs individually or SUMO3 exclusively increased either the production of APP or the secretion of Aβ or both. Furthermore, sumoylation was suggested to have an indirect role through non-covalent interactions (Dorval et al., 2007; Yun et al., 2013).

Toxic neurofibrillary tangle-forming Tau is also subjected to sumoylation, which appears to inhibit its Ubiquitin-mediated degradation (Dorval and Fraser, 2006; Luo et al., 2014). SUMO is preferentially conjugated to the soluble pool of Tau, however the physiological role of this modification requires further investigation (Dorval and Fraser, 2006; Lee et al., 2013).

These examples illustrate the involvement and importance of covalent modification by SUMO in proteopathic diseases, although many underlying mechanisms and physiological roles of sumoylation remain to be elucidated.

1.5. Identification of SUMO substrates

In general, the identity of SUMO conjugates is inferred either on a scale of a single protein using substrate-specific methods, or *en masse* by large-scale proteomic analyses of sumoylation. Both approaches are hindered by the low abundance of targets and limited occupancy of sumoylation (Hay, 2005; Wohlschlegel et al., 2006; Tatham et al., 2009). To improve the detection of modified substrates, SUMO is often N-terminally fused to an affinity tag, and proteins conjugated to the exogenously expressed SUMO are purified by affinity chromatography. Various protein tags have been successfully used for the enrichment of mammalian sumoylated substrates, including polyhistidine (Rodriguez et al., 1999; Vertegaal et al., 2004; Vertegaal et al., 2006), Human influenza virus hemagglutinin (HA) (Zhao et al., 2004; Blomster et al., 2009), tandem affinity purification (TAP) tag (Golebiowski et al., 2009; Tatham et al., 2011; Domingues et al., 2015), FLAG (Schimmel et al., 2014), and Strep-tag (Bursomanno et al., 2015). In some instances, the best strategy has been the combined utilisation of immobilised metal- and immunoaffinity chromatography (Schou et al., 2014). In addition to cultured cell lysates, affinity purification techniques have also been used to enrich and identify SUMO conjugates from knock-in mice (Tirard et al., 2012).

Hitherto, only a limited number of studies have accomplished the high-throughput identification of endogenous SUMO substrates. These methods still require the enrichment of sumoylated proteins and thus include immunoaffinity chromatography techniques using either SUMO paralog-specific antibodies (Matafora et al., 2009; Becker et al., 2013; Cubenas-Potts et al., 2015) or a fragment of RNF4 that binds to poly-SUMO chains (Bruderer et al., 2011).

The modification status of a specific putative target protein is often determined by Western blot analyses (Desterro et al., 1998; Rodriguez et al., 1999; Muller et al., 2000). High-throughput identification of sumoylated substrates is facilitated by mass spectrometry (MS)-based proteomics. The extensive global investigations have altogether revealed hundreds of putative SUMO-modified proteins from human cells cultured either unperturbed (Vertegaal et al., 2006; Becker et al., 2013) or in response to various cellular stresses, such as DNA damage (Bursomanno et al., 2015; Hendriks et al., 2015b), proteasome inhibition (Schimmel et al., 2008; Matafora et al., 2009; Tatham et al., 2011), exposure to heat (Golebiowski et al., 2009; Bruderer et al., 2011), or viral infection (Domingues et al., 2015; Sloan et al., 2015).

One of the key challenges of the proteomic studies of sumoylation has been the separation of genuine SUMO-modified substrates from the co-purified "contaminant" proteins that are identified in the same experiment (Golebiowski et al., 2009). Lack of consideration for this technical issue could lead to a false conclusion and an inclusion of many contaminants to a list of SUMO substrates. As the direct evidence of sumoylation is often absent, the target proteins are inferred indirectly by a process of elimination (Vertegaal et al., 2006; Golebiowski et al., 2009; Bruderer et al., 2011). Direct evidence of modification would be provided by the identification of sequences containing the modified lysine(s) attached to the C-terminal peptide of the SUMO itself. Methods for the high-throughput mapping of SUMO modification sites, however, are lacking. In many investigations, the lysines most likely to be targeted for SUMO conjugation are systematically mutated to chemically similar non-modifiable residues, and genuine substrates are inferred by assessing the ability of these mutant proteins to serve as substrates of sumoylation (Desterro et al., 1998; Rodriguez et al.,

1999; Muller et al., 2000). However, these mutational studies identify the residues required for modification, but do not provide direct evidence at the mutated site. Several independent research groups have also acknowledged the benefits of MS-based proteomics for the high-throughput, unbiased, and direct identification of SUMO modification sites, but the developed strategies encountered several technical challenges (discussed in Chapter 3.1.1) (Cooper et al., 2005; Blomster et al., 2010; Matic et al., 2010; Galisson et al., 2011; Lamoliatte et al., 2013). Therefore, prior to the launch of this PhD project, the largest site-specific study of protein sumoylation determined only 103 sumoylated lysines (Matic et al., 2010), which was far fewer than expected based on the number of putative SUMO-modified proteins identified thus far.

2.

Materials and methods

2.1. Cell culture procedures

2.1.1. Cultured cell lines

Cell lines used in this study are listed in Table 2.1.

Table 2.1 | Cell lines.

Name	Description	Source	Cat. no.
HEK293 N3S	Human embryonic kidney 293 suspension cells	Sigma-Aldrich	92052131
HEK293$^{6His\text{-}SUMO2}$ N3S	6His-SUMO2-expressing human embryonic kidney 293 suspension cells	Hay lab stock	
HEK293$^{6His\text{-}SUMO2\text{-}T90K}$ N3S	6His-SUMO2^{T90K}-expressing human embryonic kidney 293 suspension cells	This work	
HeLa	Human cervix epitheloid carcinoma cells	Hay lab stock	

2.1.2. Generation of HEK293$^{6His\text{-}SUMO2\text{-}T90K}$ N3S cells

HEK293$^{6His\text{-}SUMO2\text{-}T90K}$ N3S cell line was generated by Adel F. M. Ibrahim (University of Dundee). A bicistronic pEFIRESpuro-6His-SUMO2^{T90K} plasmid was created by replacing the coding sequence of enhanced Yellow fluorescent protein (eYFP) of a vector pEFIRESpuro-eYFP-C1 with a polymerase chain reaction (PCR)-produced 6His-SUMO2 fusion sequence (Hobbs et al., 1998). NheI restriction enzyme site was coded between the sequences for SUMO2 and the 6His-tag. T90K mutation was introduced into the coding sequence of 6His-SUMO2 by site-directed mutagenesis, and

the region encoding 6His-SUMO2^{T90K} was submitted for sequencing to the DNA Sequencing facility at the University of Dundee.

HEK293 N3S cells cultured in suspension were transfected with a plasmid pEFIRESpuro-6His-SUMO2^{T90K} using Lipofectamine 2000 (Thermo Fisher Scientific, cat. no. 11668019) and the selection of stably-transfected cells was performed by supplementation with 2 μg/ml puromycin. The generated stable HEK293$^{6His\text{-}SUMO2\text{-}T90K}$ N3S cell line was maintained in a puromycin-containing (1 μg/ml) culture medium at 37 °C in 5 % CO_2.

The amino acid sequence of the stably-expressed 6His-SUMO2^{T90K} is:

MHHHHHHASMSEEKPKEGVKTENDHINLKVAGQDGSVVQFKIKRHTPLSKLMKA
YCERQGLSMRQIRFRFDGQPINETDTPAQLEMEDEDTIDVFQQQKGG

2.1.3. Cell culture conditions for label-free experiments

Adherent HEK293$^{6His\text{-}SUMO2\text{-}T90K}$ N3S or HEK293$^{6His\text{-}SUMO2}$ N3S cells were cultured in Dulbecco's modified Eagle medium (DMEM; Thermo Fisher Scientific, cat. no. 61965026) supplemented with 10 % fetal bovine serum (FBS; Biosera, cat. no. FB-1090), 1 μg/ml puromycin (Melford, cat. no. P0121), and 50 U/ml penicillin with 50 μg/ml streptomycin (100 U/ml Pen-Strep; Thermo Fisher Scientific, cat. no. 15140122). Adherent HeLa or HEK293 N3S cells were cultured in DMEM supplemented with 10 % FBS and 100 U/ml Pen-Strep, and maintained at 37 °C in 5 % CO_2.

For the label-free experiments using SUMO^{KGG} workflow, HEK293$^{6His\text{-}SUMO2\text{-}T90K}$ N3S cells were grown on five 175 cm^2 dishes to about 90 % confluency prior to their transfer

into Minimum Essential medium Eagle (SMEM; Sigma-Aldrich, cat. no. M8167) supplemented with 10 % FBS, 1 μg/ml puromycin, 2 mM L-glutamine (Thermo Fisher Scientific, cat. no. 25030024), and 100 U/ml Pen-Strep. Approximately 375 ml of suspension culture corresponding to ~1.75×10^8 cells was cultured per experiment, stimulated by heat shock at 43 °C for 30 minutes, harvested by centrifugation, and washed with cold 1 × Dulbecco's phosphate buffered saline (DPBS; Thermo Fisher Scientific, cat. no. 14190094).

The time course experiments of the heat shock response were performed with adherent cultures of HEK293$^{6His\text{-}SUMO2\text{-}T90K}$ cells incubated at 43 °C for 0, 15, 30, 60, or 120 minutes. Cells were scraped into the culture medium, harvested by centrifugation and washed twice with cold 1 × DPBS.

Immunofluorescence microscopy experiments were performed with adherent HeLa cells cultured in two 24-well plates, and one of the plates was incubated at 43 °C for 30 minutes. Unperturbed and heat-stressed HeLa cells were washed twice with 1 × DPBS at either 37 °C or 43 °C, respectively.

2.1.4. Cell culture conditions for SILAC-based experiments

For quantitive proteomic experiments, the adherent HEK293$^{6His\text{-}SUMO2}$ N3S or HEK293$^{6His\text{-}SUMO2\text{-}T90K}$ N3S cells were cultured in DMEM lacking L-lysine, L-arginine and L-glutamine (Biosera; cat. no. SM-D0550), and supplemented with 10 % dialysed FBS, 1 μg/ml puromycin, 100 U/ml Pen-Strep, 2 mM L-glutamine, and either natural [Sigma-Aldrich, cat. nos. L8662 ($^{12}C_6$, $^{14}N_2$; K0) and A8094 ($^{12}C_6$, $^{14}N_4$; R0)] or heavy stable isotope-containing [Cambridge Isotope Laboratories, cat. nos. CNLM-291-H ($^{13}C_6$, $^{15}N_2$;

K8) and CNLM-539-H ($^{13}C_6$, $^{15}N_4$; R10)] 0.8 mM L-lysine and 0.14 mM L-arginine. Five 175 cm^2 dishes of either light- or heavy-labelled cells at ~90 % confluency were transferred into SMEM lacking L-lysine, L-arginine and L-glutamine (Biosera; cat. no. SM-M0538), and supplemented with 10 % dialysed FBS, 1 μg/ml puromycin, 100 U/ml Pen-Strep, 2 mM L-glutamine, and either natural or heavy stable isotope-containing 0.8 mM L-lysine and 0.14 mM L-arginine, respectively. Approximately 0.5–1.0 L of suspension culture corresponding to ~2.35–4.70 × 10^8 cells was cultured per experimental condition, and either heat-stressed at 43 °C for 30 minutes, treated by supplementation with GRE010 to 20 μM for 8 hours, or left unperturbed. Cells were harvested by centrifugation and washed twice with cold 1 × DPBS.

Proteomic experiments involving subcellular fractionation were performed with HEK293$^{6His\text{-}SUMO2\text{-}T90K}$ cells cultured as above, and ~3 L of suspension culture was harvested per experimental condition.

2.1.5. Cell proliferation assay

The growth rates of parental or 6His-SUMO2^{T90K}-expressing HEK293 N3S cells were determined in collaboration with Adel. F. M. Ibrahim (University of Dundee). Adherent cultures of HEK293 N3S or HEK293$^{6His\text{-}SUMO2\text{-}T90K}$ N3S cells were subcultured at a low density (<1000 cells/well) into 96-well plates (Greiner Bio-one, cat. no. 655090). The 6His-SUMO2^{T90K}-expressing cells were analysed twice, either with or without the supplementation with 1 μg/ml puromycin. Both cell lines were subcultured in 32 replicates per time point, and after 2, 3, 4, or 5 days of incubation at 37 °C, the cells were fixed with 4 % formaldehyde followed by staining with 0.1 μg/ml 4',6-diamidino-2-phenylindole (DAPI; Sigma-Aldrich, cat. no. D9542). The images were acquired with

high-content IN Cell Analyzer 2000 imaging system (GE Healthcare Life Sciences), and the cells were counted using the IN Cell Investigator Developer Toolbox 1.9.1, build 2206 (GE Healthcare Life Sciences).

2.1.6. Cell lysis without subcellular fractionation

Cells for proteomic experiments, which did not require subcellular fractionation, were collected by centrifugation, lysed in fresh cell lysis buffer containing 6 M guanidine hydrochloride (Gu-HCl), 100 mM sodium phosphate buffer pH 8.0 [5.3 % (vol/vol) NaH_2PO_4 in Na_2HPO_4], 10 mM Tris-HCl pH 8.0, 10 mM imidazole, and 5 mM 2-mercaptoethanol. Lysis buffer was added to cell pellets in a ratio of 5:1 (vol/wt). DNA was sheared by short pulses of sonication, insoluble material was removed by centrifugation and filtration through sterile Minisart NML syringe filters with 0.2 μm pore size (Sartorius, cat. no. 16534), and protein concentration was estimated using Pierce bicinchoninic acid (BCA) protein assay (Thermo Fisher Scientific, cat. nos. 23223 and 23224). Cells cultured for Western blot analysis-based experiments were lysed in 4 % sodium dodecyl sulfate (SDS), 100 mM Tris-HCl pH 7.5 and incubated at 95 °C for 1–2 minutes.

2.1.7. Cell lysis with subcellular fractionation

Cells harvested for the proteomic analyses that involved subcellular fractionation (SF) were resuspended in cold SF buffer containing 10 mM HEPES [4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid] pH 7.9, 1.5 mM $MgCl_2$, 10 mM KCl, 0.08 % NP-40, 100 mM 2-chloroacetamide, and cOmplete protease inhibitor cocktail (Sigma-Aldrich, cat. nos. 11836170001 and 11873580001), and incubated on ice for 15 minutes followed

by 30-minute rotation at 4 °C. Nuclei were separated from cytoplasmic material by centrifugation at 2000 g for ~5 minutes, washed with cold SF buffer, and subcellular compartments and protein structures in each fraction were disrupted by supplementation with 6 M Gu-HCl, 10 mM imidazole. DNA was sheared by sonication, insoluble material was removed by centrifugation and filtration through sterile Minisart NML syringe filters with 0.2 μm pore size, and protein concentration was estimated using Pierce BCA protein assay. Cells cultured for Western blot analysis-based experiments were prepared as above, but 6 M Gu-HCl was replaced with 2 % SDS.

2.2. Protein techniques

2.2.1. *In vitro* SUMO conjugation, deconjugation and processing

In vitro assays were performed by Ellis G. Jaffray (University of Dundee). All reactions were buffered in 50 mM Tris-HCl pH 7.5, and samples were analysed by SDS-polyacrylamide gel electrophoresis (PAGE).

Conjugation assays contained 5 mM dithiothreitol (DTT), 5 mM $MgCl_2$, 2 mM ATP, 110 nM SAE1/SAE2, 0.5–2 μM Ubc9, ~10 μM substrate protein and a range of mature recombinant SUMO: 0, 40, 80, or 200 μM. The reactions were incubated at 37 °C for 4 hours.

Deconjugation assays were prepared as above using 200 μM SUMO followed by a supplementation with SENP1 to 10 nM, and the reactions were monitored at 0, 0.5, 1, 2.5, 5 and 10 minutes at 20 °C.

Precursor SUMO processing assays contained 150 mM NaCl, 0.5 mM Tris(2-carboxyethyl)phosphine (TCEP), 600 μM SUMO2, and 100 nM or 200 nM of recombinant catalytic domain of SENP1 or SENP2, respectively (Shen et al., 2006). The reactions were incubated at 20 °C for 0, 5, 10, 20, 30, 60, 90, 120, 180, 240 and 960 minutes.

The amino acid sequences of the precursor or mature recombinant SUMO2 and $SUMO2^{T90K}$ proteins used in these assays are:

- Precursor SUMO2:

GSEEKPKEGVKTENDHINLKVAGQDGSVVQFKIKRHTPLSKLMKAYCERQGLSMRQ
IRFRFDGQPINETDTPAQLEMEDEDTIDVFQQQTGGVPESSLAGHSF

- Mature SUMO2:

GSEEKPKEGVKTENDHINLKVAGQDGSVVQFKIKRHTPLSKLMKAYCERQGLSMRQ
IRFRFDGQPINETDTPAQLEMEDEDTIDVFQQQTGG

- Precursor SUMO2^{T90K}:

GSEEKPKEGVKTENDHINLKVAGQDGSVVQFKIKRHTPLSKLMKAYCERQGLSMRQ
IRFRFDGQPINETDTPAQLEMEDEDTIDVFQQQKGGVPESSLAGHSF

- Mature SUMO2^{T90K}:

GSEEKPKEGVKTENDHINLKVAGQDGSVVQFKIKRHTPLSKLMKAYCERQGLSMRQ
IRFRFDGQPINETDTPAQLEMEDEDTIDVFQQQKGG

2.2.2. SDS-PAGE, Western blot analyses, and Coomassie brilliant blue staining

SDS-PAGE was performed using either 10 % or 4–12 % NuPAGE Novex Bis-Tris precast protein gels (Thermo Fisher Scientific, cat. nos. NP0301–NP0303, NP0321–NP0323, or WG1402–WG1403) in MOPS [3-(N-morpholino)propanesulfonic acid] or MES [2-(N-morpholino)ethanesulfonic acid] SDS running buffer, and macromolecules were separated at a constant voltage of 150 V.

Western blot images were prepared using Immobilon-P polyvinylidene difluoride (PVDF) transfer membrane (Merck Millipore, cat. no. IPVH00010), Pierce enhanced chemiluminescence (ECL) Western blotting substrate (Thermo Fisher Scientific, cat. no 32106), and antibodies specified in Table 2.2.

Coomassie brilliant blue staining of protein gels or Western blot PVDF membranes was performed with 0.18 % (wt/vol) Coomassie brilliant blue R-250 in 9.1 % (vol/vol) acetic acid, 45.5 % (vol/vol) ethanol, followed by destaining with 7 % acetic acid in 40 % ethanol and short-term storage in 7 % acetic acid, 5 % ethanol.

2.2.3. Protein concentration estimation

Cell lysate proteins diluted either two-, five- or tenfold into water were mixed in triplicate with a solution of Pierce BCA protein assay reagents in a ratio of ~1:30 (vol/vol). A range of concentrations (0.05, 0.1, 0.25, 0.5, and 1.0 mg/ml) of bovine serum albumin (BSA; Sigma-Aldrich, cat. no. A7906) was used to generate the calibration curve. Mixtures were incubated at 60 °C for 25 minutes, at 4 °C for 30 minutes, and protein concentration in each cell lysate was estimated by UV-visible spectrophotometry at ~570 nm.

2.2.4. Antibodies

Antibodies used in this study are listed in Table 2.2.

Table 2.2 | Antibodies.

Name	Species	Source	Catalog no.	Appl.	Dilution
6×His	Mouse	Clontech	631212	WB	1:3000
SUMO2	Rabbit	Thermo Fisher Scientific	91-5100 (discontinued)	WB	1:1500
SUMO-2/3 (18H8)	Rabbit	Cell Signaling Technology	#4971S	IF	1:100
HSF1	Rabbit	Cell Signaling Technology	#4356S	IF	1:500
Tubulin α (YOL1/34)	Mouse	Thermo Fisher Scientific	MA1-80189	WB	1:5000
Lamin A/C (636)	Mouse	Santa Cruz Biotechnology	sc-7292	WB	1:1000 or 1:500
Anti-rabbit-peroxidase secondary antibody	Goat	Sigma-Aldrich	A6154	WB	1:3000
Anti-mouse-peroxidase secondary antibody	Goat	Sigma-Aldrich	A9917	WB	1:3000
Anti-rabbit secondary antibody, Alexa Fluor 488	Goat	Thermo Fisher Scientific	A11034	IF	1:500
PTMScan Ubiquitin remnant motif (K-ε-GG) kit	Rabbit	Cell Signaling Technology	#5562	IAP	

WB, Western blot; IF, immunofluorescence microscopy; IAP, immunoaffinity purification

2.2.5. Cross-linking of K-ε-GG-specific antibody

Chemical cross-linking of diGly-modified lysine (K-ε-GG)-specific antibody (Table 2.2) was performed with a fresh solution of bis(sulfosuccinimidyl)suberate (BS3; Thermo Fisher Scientific, cat. no. 21585). The antibody-conjugated protein A agarose beads were washed twice with 20 resin volumes of conjugation buffer containing 150 mM NaCl in 20 mM sodium phosphate buffer pH 9.0 [1.55 % (vol/vol) NaH_2PO_4 in

Na_2HPO_4], resuspended in 25 resin volumes of 5 mM BS3 in conjugation buffer, and rotated at room temperature (~22 °C) for 30 minutes. The cross-linking reaction was quenched by supplementation with Tris-HCl pH 7.5 to 50 mM and cross-linked beads were washed three times with 20 resin volumes of cold immunoaffinity purification (IAP) buffer (50 mM MOPS-NaOH pH 7.2, 10 mM Na_2HPO_4, 50 mM NaCl), resuspended in cold IAP buffer in a final concentration of 50 % (vol/vol), and stored at 4 °C.

2.2.6. Protein precipitation

Up to 200 μg of proteins were supplemented with cold 10 % (vol/vol) trichloroacetic acid (TCA; Sigma-Aldrich, cat. no. T6508) in a final volume of 200 μl, and incubated on ice for one hour. Proteins were collected by centrifugation, washed twice with cold ethanol and resuspended in buffer solutions compatible with specific downstream analyses.

2.3. Sample preparation for MS

2.3.1. Nickel affinity chromatography

Nickel affinity purification of 6His-SUMO2^{T90K} conjugates was performed according to a protocol adapted from (Tatham et al., 2011). Cell lysate proteins were mixed with pre-equilibrated nickel-nitriloacetic acid (Ni^{2+}-NTA) agarose resin (Qiagen, cat. no. 30210) in a ratio of 200:1 (wt/vol), and the slurry was mixed overnight (14–18 hours) at 4 °C. Beads were collected into an empty spin chromatography column (Bio-Rad, cat. no. 7326204) and washed with five resin volumes of cell lysis buffer, ten resin volumes of wash buffer pH 8.0 (8 M urea, 100 mM sodium phosphate buffer pH 8.0, 10 mM Tris-HCl pH 8.0, 10 mM imidazole, 5 mM 2-mercaptoethanol), ten resin volumes of wash buffer pH 6.3 (8 M urea, 100 mM sodium phosphate buffer pH 6.3 [77.5 % (vol/vol) NaH_2PO_4 in Na_2HPO_4], 10 mM Tris-HCl pH 8.0, 10 mM imidazole, 5 mM 2-mercaptoethanol), and ten resin volumes of wash buffer pH 8.0. Conjugates were eluted in three sequencial steps with two resin volumes of elution buffer (200 mM imidazole in wash buffer pH 8.0) and protein concentration was estimated by UV-visible spectrophotometry at 280 nm.

2.3.2. Filter-aided sample preparation and protein digestion

Protein digestion was performed on an ultrafiltration spin column with a 30 kDa nominal molecular weight cutoff limit (Sartorius, cat. no. VN01H22) according to a protocol adapted from (Manza et al., 2005; Wisniewski et al., 2009b). Denatured proteins were concentrated on the device, washed twice with 8 M urea, 100 mM Tris-HCl pH 7.5, and treated with 50 mM 2-chloroacetamide in 8 M urea, 100 mM

Tris-HCl pH 7.5 at room temperature for 20 minutes in the dark. Samples were washed twice with IAP buffer and digested overnight with Lysyl endopeptidase (Lys-C; Wako, cat. no. 129-02541) in IAP buffer at 37 °C [enzyme-to-protein ratio 1:50 (wt/wt)]. Peptides were collected and the device was washed with IAP buffer to increase the the yield of Lys-C digested peptides. High-molecular-weight peptides retained on the 30 kDa ultrafiltration spin column were subsequently digested with endoproteinase Glu-C (Sigma-Aldrich, cat. no. 11047817001) overnight in IAP buffer at 25 °C [enzyme-to-protein ratio 1:100 (wt/wt)], and after the collection of peptides, the device was washed with IAP buffer.

Experiments evaluating the specificity of endoproteinases were performed as above, but digestion with Lys-C was performed under two conditions, either 25 mM ammonium bicarbonate (NH_4HCO_3) or IAP buffer. After Lys-C digestion, the samples were supplemented with either Glu-C or trypsin (Promega, cat. no. V5280) at enzyme-to-protein ratio of 1:50 (wt/wt) and digested overnight at either room temperature or 37 °C, respectively.

2.3.3. In solution digestion of proteins

Complete proteome analyses were performed with cell lysate proteins dissolved in 6 M urea, 2 M thiourea after TCA-based precipitation. The proteins were treated with 10 mM DTT for 1 hour and 50 mM 2-chloroacetamide for 1.5 hours at room temperature in the dark, prior to 2.5-fold dilution into 50 mM ammonium bicarbonate and digestion with Lys-C at enzyme-to-protein ratio 1:50 (wt/wt) at room temperature for 4.5 hours. Lys-C digested samples were then equally divided in two, and one of the samples was diluted twofold into 50 mM ammonium bicarbonate followed by

supplementation with trypsin at enzyme-to-protein ratio 1:50 (wt/wt). Both samples were digested overnight at room temperature. When stated, resulting peptide mixtures were fractionated into six fractions based on the pH of the solution (pH 11.0, 8.0, 6.0, 5.0, 4.0, and 3.0) used to elute the peptides from a pipette tip-based Empore anion exchanger (Agilent Technologies, cat. no. 12145012) according to a protocol described in (Wisniewski et al., 2009a).

2.3.4. Immunoaffinity purification of diGly-Lys-containing peptides

The enrichment of diGly-Lys-containing peptides was performed according to a protocol adapted from (Udeshi et al., 2013a). Peptides in IAP buffer were supplemented with ~18.75 μg of K-ε-GG-specific antibody cross-linked to 3 μl of protein A agarose resin, and rotated overnight at 4 °C. Agarose beads were washed twice with cold IAP buffer and peptides were eluted in two sequencial steps with 0.15 % (vol/vol) trifluoroacetic acid (TFA).

2.3.5. Desalting of peptides

Prior to MS analyses, all peptide samples were concentrated and desalted on Empore C18 (Agilent Technologies, cat. no. 12145004) Stop and Go Extraction Tips (StageTips) according to a protocol adapted from (Rappsilber et al., 2007). Self-made StageTips stacked with three layers of C18-bonded silica extraction disks were washed with methanol, conditioned with 80 % (vol/vol) acetonitrile, 0.1 % (vol/vol) TFA, and equilibrated with 0.1 % (vol/vol) TFA. Up to 15 μg of peptides in 0.1–0.15 % TFA were

loaded onto StageTips, washed with 0.1 % TFA, extracted with 80 % acetonitrile, 0.1 % TFA, and vacuum concentrated at 30 °C prior to the reconstitution in 0.1 % TFA.

2.3.6. SUMO^{KGG} workflow

A detailed protocol of SUMO^{KGG} workflow is provided in Appendix 6.2 (Tammsalu et al., 2015) and summarised in Figure 3.8.

2.4. MS procedures

2.4.1. Liquid chromatography (LC)-tandem mass spectrometry (MSMS)

All desalted peptide samples were analysed using EASY-nLC 1000 nano-flow UHPLC system, EASY-Spray ion source and Q Exactive hybrid quadrupole-Orbitrap mass spectrometer (all Thermo Fisher Scientific). Peptides were loaded onto 2 cm Acclaim PepMap 100 C18 nanoViper pre-column (75 μm inner diameter; 3 μm particles; 100 Å pore size) at a constant pressure of 800 bar and separated using 50 cm EASY-Spray PepMap RSLC C18 analytical column (75 μm inner diameter; 2 μm particles; 100 Å pore size) maintained at 45 °C.

DiGly-Lys-enriched samples were analysed at least twice. Exploratory analysis using standard MS settings was in general performed with 10 % of the sample and peptides were separated with 60-minute linear gradient of 5–22 % (vol/vol) acetonitrile in 0.1 % (vol/vol) formic acid at a flow rate of 250 nl/min, followed by a 12-minute linear increase of acetonitrile to 40 % (vol/vol). Total length of the gradient including column washout and re-equilibration was 90 minutes. Comprehensive peptide analyses were performed using either identical conditions or a 120-minute linear gradient of 5–22 % (vol/vol) acetonitrile in 0.1 % (vol/vol) formic acid at a flow rate of 250 nl/min, with a subsequent 15-minute linear increase of acetonitrile to 40 % (vol/vol). Thus, the overall lengths of the gradients during comprehensive analyses were either 90 or 150 minutes, respectively. Peptides corresponding to the complete human proteome were in general analysed with 240-minute linear gradient of acetonitrile in 0.1 % (vol/vol) formic acid

at a flow rate of 250 nl/min, with the majority of peptides eluting during a 220-minute acetonitrile window from 5 % to 50 % (vol/vol).

Peptides eluting from the liquid chromatography (LC) column were charged using electrospray ionisation and MS data was acquired online in a profile spectrum data format. Full MS scan covered a mass range of mass-to-charge ratio (m/z) 300−1800 or 300−1600 during standard or comprehensive peptide analyses, respectively. Target value was set to 1 000 000 ions with a maximum injection time (IT) of 20 ms and full MS was acquired at a mass resolution of 70 000 at m/z 200. Data dependent MSMS scan was initiated if the intensity of a mass peak reached a minimum of 20 000 ions. During standard LC-MSMS analyses, up to 10 most abundant ions (Top 10) were selected using 2 Th mass isolation range when centered at the parent ion of interest. For comprehensive analyses, either up to 5 (Top 5) or the most abundant ion exclusively (Top 1) were picked for MSMS. Selection of molecules with peptide-like isotopic distribution was preferred. Target value for MSMS scan was set to 500 000 ions with a maximum IT of 60 ms and resolution of 17 500 at m/z 200 for standard, or maximum IT of 1000 ms (Top 1) or 300 ms (Top 5), and a resolution 35 000 at m/z 200 for comprehensive peptide analyses. Precursor ions were fragmented by higher energy collisional dissociation (HCD) using normalised collision energy of 30 and fixed first mass was set to m/z 100. Precursor ions with undetermined, single, or high (>8) charge state were rejected. Ions triggering a data-dependent MSMS scan were placed on the dynamic exclusion list for 40 s (standard analyses) or 60 s (comprehensive analyses) and isotope exclusion was enabled.

2.4.2. Analysis of raw MS files

Raw mass spectrometric data files were processed with MaxQuant software (versions 1.3.0.5 or 1.5.2.8) (Cox and Mann, 2008) and peak lists were searched with an integrated Andromeda search engine (Cox et al., 2011) against an entire human UniprotKB proteome containing canonical and isoform sequences (The UniProt Consortium, 2015). Raw files were divided into parameter groups based on the specificity of proteolysis applied during sample preparation. Hydrolysis of peptide bonds C-terminal to Lys, or Lys and Arg with a maximum of three missed cleavages was allowed for peptides processed with Lys-C, or Lys-C and trypsin, respectively. Samples acquired after an additional Glu-C digestion were analysed with enzyme specificity set to C-terminal to Lys, Glu and Asp with a maximum of five missed cleavages. Carbamidomethylation of cysteine residues was specified as a fixed modification and oxidation of methionines, acetylation of protein N-termini, and where applicable, Gly-Gly adduct on internal lysine residues were selected as variable modifications. When required, phosphorylation of Ser, Thr and Tyr was also included as a variable modification. Maximum peptide mass of 10 000 Da was allowed, multiplicity was set to 1 and re-quantify option was disabled. Decoy sequence database was generated using Lys as a special amino acid. Default values were chosen for the rest of the parameters. For SILAC-based samples, multiplicity was set to two, and K8R10 were selected as heavy-labelled counterparts. Match between runs option was enabled with a match time window of 1 min and alignment time window of 20 min. All data sets were filtered by posterior error probability to achieve a false discovery rate of 1 % at protein, peptide and site level.

2.5. Microscopy techniques

2.5.1. Immunofluorescence microscopy

Immunofluorescence microscopy experiments were performed in collaboration with Ellis G. Jaffray and Federico Pelisch (both University of Dundee). Unperturbed or heat-stressed HeLa cells in 24-well plates were fixed with 3.8 % formaldehyde in phosphate buffered saline (PBS), permeabilised with 0.2 % Triton X-100 in PBS, blocked with 5 % BSA in 0.1 % Tween 20, PBS, incubated with antibodies specified in Table 2.2, and stained with 0.1 μg/ml DAPI. The samples were mounted on microscope slides (VWR, cat. no. 631-0116) and images were acquired using a widefield DeltaVision Elite imaging system (GE Healthcare Life Sciences), an Olympus IX71 inverted microscope with a 60 ×/1.42 Plan Apochromat oil immersion lens (Olympus), a CoolSNAP HQ camera (Photometrics), and softWoRx software. Non-deconvolved images were processed using ImageJ 1.50b (Schneider et al., 2012), maintaining all parameters when comparing unperturbed or heat-stressed cells.

2.5.2. Light microscopy

Photographs of adherent HEK293 N3S or HEK293$^{6His\text{-}SUMO2\text{-}T90K}$ N3S cells cultured in 75 cm^2 flask were taken using an Axiovert 40 CFL inverted microscope, an LD A-Plan 20 ×/0.3 Ph1 objective (both Carl Zeiss Microscopy), and a Panasonic Lumix DMC-G3 camera.

2.6. Bioinformatic analyses

2.6.1. Alignments

Sequence alignments were generated with a Clustal Omega multiple sequence alignment tool (Sievers et al., 2011) using Jalview, version 2 editor (Waterhouse et al., 2009).

Structural alingments were created with a protein structure comparison service PDBeFold, version 2.59 at European Bioinformatics Institute (http://www.ebi.ac.uk/msd-srv/ssm) (Krissinel and Henrick, 2004).

2.6.2. Sequence analyses

Primary structure motif analyses were performed with pLogo (O'Shea et al., 2013). Sequence windows of 13 amino acids centered at the modified lysine were used as an input, and N- or C-terminal sequences not covering the 13 residue window were omitted from the output. Residues were scaled relative to their Bonferroni-corrected statistical significance using human proteome as a background data set (645 531 lysine residues).

2.6.3. Pathway analyses

Protein function annotation and network analyses were generated using Qiagen's Ingenuity Pathway Analysis (IPA; www.qiagen.com/ingenuity).

Cellular component-based annotations were retrieved from Gene Ontology database (released in August 2016) using PANTHER Classification System (The Gene Ontology Consortium, 2015; Mi et al., 2016) and analysed with PANTHER overrepresention tool (released in July 2016) using Bonferroni-correction for multiple testing.

2.6.4. Protein secondary structure prediction

Secondary structure predictions were performed in collaboration with Christian Cole (University of Dundee). All modified peptides were expanded to their whole protein "parent" and submitted to JPred4 for protein secondary structure prediction (Drozdetskiy et al., 2015). JPred4 had length and time limits for making predictions, so proteins longer than 800 or shorter than 20 amino acids, or predictions taking longer than one hour were omitted. DisEMBL disorder prediction (Linding et al., 2003) was performed via JABAWS (Troshin et al., 2011). The central lysine residue for each peptide was then interrogated in its parent protein for its secondary structure and disorder prediction, and assigned as being either in α-helical, β-strand or coil, and disordered or non-disordered region. The counts for each of the five states were summed and compared between experimental conditions. As the JPred4-predicted secondary structures can only be one of three states, the comparison was performed at the level of proportion of each state between experimental conditions. The Wald method was used to determine the 95 % confidence interval of the proportions of each state in each data set.

2.6.5. Statistical computing

Statistical analyses were performed and graphics were generated using either GraphPad Prism version 6.0e for Mac OS X (GraphPad Software, www.graphpad.com) or R programming language (R Core Team, 2015) via RStudio interface (version 0.99.484) (RStudio Team, 2015) and implementing R packages ”stats” (R Core Team, 2015) and ”ggplot2” (Wickham, 2009).

3.

Development of a mass spectrometry-based method for the system-wide identification of SUMO modification sites

3.1. Introduction

The aim of this project was to establish a high-throughput approach for the global identification of SUMO modification sites from a diverse collection of complex protein mixtures. As a consequence of the remarkable advances in instrumentation and methodology, mass spectrometry (MS)-based proteomics has become the technique of choice for comprehensive identification and quantification of entire proteomes (Cox and Mann, 2007; Wilhelm et al., 2014). With current MS-based proteomics tools, high-accuracy qualitative and quantitative information on thousands of proteins can be rapidly collected with a single study (Nilsson et al., 2010). Moreover, numerous robust MS-based techniques have also been developed for the large-scale detection, mapping and quantification of protein posttranslational modifications (PTMs) (Cox and Mann, 2007). For these reasons, the workflow for the global identification of SUMO modification sites was also designed to utilise the advanced equipment and methodology of mass spectrometry-based proteomics.

In this chapter, the development of an experimental strategy that enables high-throughput identification of sumoylated proteins and allows mapping of exact sites of modification, is described. The importance of each stage of the workflow is evaluated, and if applicable, compared with alternative techniques that can be utilised for a similar purpose. In addition, the eligibility of individual cell lines, reagents and equipment is examined. A number of small-scale studies on SUMO2 modification sites in human kidney cells are performed to investigate and validate the robustness, reproducibility and suitability of the workflow.

3.1.1. Past limitations of mapping the modification sites of the sumoylated proteome

The reversible nature of sumoylation is achieved by a unique group of enzymes that can attach or remove SUMO from substrate proteins. The activity and specificity of deconjugating enzymes, universally termed as SUMO proteases, is largely influenced by their subcellular localisation (Hickey et al., 2012). The spatial separation of these isopeptidases from their substrates, however, is lost upon cell disruption. Therefore, if SUMO-specific proteases are not rapidly inactivated, a general decrease in protein sumoylation is observed and consequently, the identification of target proteins and sites is hampered (Vertegaal et al., 2006).

Many studies focusing on individual SUMO substrates have shown that the steady-state stoichiometry of protein sumoylation is very low (< 1 %) (Hay, 2005; Wohlschlegel et al., 2006; Tatham et al., 2009). The identification of such low abundance target proteins by liquid chromatography (LC)-tandem mass spectrometry (MSMS) is often challenging, because of the wide dynamic range of individual protein abundances in a cell lysate (Nilsson et al., 2010). As discussed in Chapter 1.5, a collection of methods have been developed that address this issue by enriching SUMO-modified molecules from the rest of the proteome. However, none of them provide direct evidence of sumoylation, as the exact location of modified lysine(s) is not provided.

Prior to the LC-MSMS analysis, proteins are generally hydrolysed into peptides. This type of approach, where MS-based detection and sequencing is performed at the peptide level, is known as bottom-up or shotgun proteomics. Although mass spectrometers are able to measure the masses of intact proteins, their sensitivity and

accuracy is much higher for low-molecular-weight molecules, and the success rate of obtaining complete sequence information is better with peptides (Steen and Mann, 2004; Cox and Mann, 2011). On the other hand, proteolysis increases the sample complexity, as proteins are generally digested into multiple peptides. Importantly, individual sumoylated molecules of a substrate protein can be modified on distinct lysine(s) and most residues are not sumoylated at any particular time. Digestion of these heterogenic samples of SUMO conjugates will thus generate complex mixtures, which are characterised by an excess of unmodified over modification-specific peptides. Therefore, analogous to the complications inherent in protein-level sumoylation analysis, MS-based identification of low abundance modified peptides without prior enrichment is challenging, because of the wide dynamic range of the peptide abundances in the arising samples.

SUMO is a large PTM, and a single sumoylation event will increase the mass of a target protein by ~11 kDa (The UniProt Consortium, 2015). Like other proteins, SUMO can be proteolytically cleaved, and the C-terminal peptide of a substrate-bound SUMO molecule will remain attached to the target lysine. In most cases of bottom-up proteomics, peptides are generated using trypsin (Steen and Mann, 2004; Tsiatsiani and Heck, 2015), which is a highly specific endoproteinase that cleaves at the C-terminal side of arginines and lysines (Olsen et al., 2004). Inspection of the primary sequences of all SUMO paralogs, however, reveals a dearth of these residues in their C-terminal regions. Consequently, the digestion of sumoylated proteins with trypsin will generate long remnants attached to target lysines. In particular, substrate-bound tryptic remnants of human SUMO1 or SUMO2/3 contain 19 or 32 amino acids, respectively, and increase the mass of a modified peptide by ~2.1 or ~3.5 kDa. These SUMO

remnant-modified peptides with two N-termini, termed "branched peptides", generate intricate MSMS fragmentation spectra and to date can only be processed using specialist software, which appear to have limited applicability to complex peptide mixtures (Pedrioli et al., 2006; Matic et al., 2008; Hsiao et al., 2009).

In this chapter an MS-based approach was developed to address all abovementioned shortcomings and thus enable the systems-wide identification of the sites of SUMO modification.

3.2. Results

3.2.1. Selection of an appropriate cell lysis solution with a compatible protein affinity tag

An important consideration for the successful isolation of SUMO-modified proteins is the formulation of the cell lysis buffer. A major complication that has to be resolved during the enrichment of sumoylated proteins is the presence of highly active SUMO-specific proteases in the cell lysate. If no action is taken, these will rapidly deconjugate SUMO from its substrates (Vertegaal et al., 2006). Although SUMO-specific cysteine isopeptidases can be chemically inhibited by sulfur (S)-alkylating agents, denaturing buffer conditions have been successfully utilised to promptly unfold and inactivate proteases, and thus prevent SUMO deconjugation (Vertegaal et al., 2006; Tatham et al., 2009; Bruderer et al., 2011). In addition, these buffer conditions can disrupt the non-covalent interactions between conjugates and residual non-sumoylated proteins, and reduce the proportion of co-purifying proteins, often referred to as "contaminants" (Vertegaal et al., 2006; Tatham et al., 2009).

The range of protein affinity tags compatible with strong denaturing agents is limited (Andersen et al., 2009). Therefore, a well-characterised hexahistidine (6His)-tag was selected to generate a version of SUMO that enables the enrichment of modified proteins under denaturing conditions by immobilised metal affinity chromatography (IMAC). As the isopeptide bond with a substrate is formed via the C-terminal glycine of SUMO, the 6His-tag was linked to the N-terminus of the protein (Kamitani et al., 1997). Previous studies have confirmed that N-terminal low-molecular-weight protein

tags do not interfere with the functionality of SUMO (Vertegaal et al., 2004; Tatham et al., 2009).

Nickel-nitriloacetic acid (Ni^{2+}-NTA) agarose matrix facilitates the purification of 6His-tagged recombinant proteins and has been used for the enrichment of SUMO-modified substrates. However, conventional detergent-based cell lysis buffers cannot be employed, as Ni^{2+}-NTA agarose is incompatible with the amount of detergents required to prevent desumoylation. Instead, high concentration of chaotropic agents, such as 8 M urea or 6 M guanidine hydrochloride (Gu-HCl) have been utilised with Ni^{2+}-NTA resin (Vertegaal et al., 2004; Tatham et al., 2009). Lysis of cells was thus performed using a Gu-HCl-based buffer, which was selected over urea because of its higher denaturing properties, chemical inertness, thermostability, and high solubility at a low temperature (Poulsen et al., 2013).

In conclusion, purification of 6His-tagged SUMO conjugates by nickel affinity chromatography under denaturing conditions solves two major complications of sumoylation proteomics:

1) Aggressive SUMO proteases are inactivated rapidly, as the cell lysis procedure is carried out using strong denaturing Gu-HCl-based buffer solutions.
2) The detection of SUMO substrates by LC-MSMS is achieved, as the enrichment of low abundance sumoylated proteins decreases the complexity of the samples.

3.2.2. 6His-SUMO2^{T90K} as a biochemical tool

To enable the system-wide identification of the exact sites of SUMO modification by LC-MSMS, three additional problems have to be addressed:

1) Derivatisation, substitution or truncation of the long SUMO remnant peptide is required to enable the mapping of sumoylation sites using conventional MS software.
2) The newly designed remnant sequence of SUMO should represent a unique mass-to-charge (m/z) signature to allow its discrimination from other PTMs.
3) The remnant sequence of SUMO should also enable the specific enrichment of low abundance modified peptides.

To address these points, a 6His-SUMO2^{T90K} construct was generated by site-directed mutagenesis (implemented by Adel F. M. Ibrahim). By replacing threonine^{90}, which precedes the conserved Gly-Gly motif at the C-terminus of SUMO2, with lysine (Figure 3.1; red), a new cleavage site for two highly specific endoproteinases, trypsin and Lys-C, was created. Considering that trypsin cleaves substrates C-terminal to lysines and arginines (Olsen et al., 2004), it cannot be utilised to generate SUMO remnants with a unique mass. In particular, three human Ubls (Ubiquitin, ISG15 and NEDD8) have an arginine in the equivalent position to T90K mutation (Figure 3.1), so all of them will be cleaved by trypsin and thus produce identical Gly-Gly remnants attached to target lysines. In contrast, Lys-C hydrolyses peptide bonds at the C-terminal side of lysines alone (Raijmakers et al., 2010). Notably, none of the other known human Ubls contain a lysine in the equivalent position to T90K (Figure 3.1). Therefore, Lys-C

IENVKAKIQDKEGIPPDQQRLIFAGKQLEDGRTLSDYNIQKESTLHLVLRLRGG	Ubiquitin
VERIKERVEEKEGIPPQQQRLIYSGKQMNDEKTAADYKILGGSVLHLVLALRGG	NEDD8
VAHLKQQVSGLEGVQDDLFWLTFEGKPLEDQLPLGEYGLKPLSTVFMNLRLRGG	ISG15
VAQVKAMIETKTGIIPETQIVTCNGKRLEDGKMMADYGIRKGNLLFLACYCIGG	FAT10
VAQIKAHVASLEGIAPEDQVVLLAGAPLEDEATLGQCGVEALTTLEVAGRMLGG	FUBI
LKERPELFIQGDSVRPGILVLINDA--DWELLGELDYQLQDQDSVLFISTLHGG	URM1
LKKLKESYCQRQGVPMNSLRFLFEGQRIADNHTPKELGMEEEDVIEVYQEQTGG	SUMO1
LSKLMKAYCERQGLSMRQIRFRFDGQPINETDTPAQLEMEDEDTIDVFQQQTGG	SUMO2/3
LSKLMKAYCERQGLSMRQIRFRFDGQPINETDTPAQLEMEDEDTIDVFQQQKGG	SUMO2^{T90K}

Figure 3.1 | Digestion of human Ubl-conjugated proteins with Lys-C produces a Gly-Gly remnant on target lysines modified solely by SUMO2^{T90K}.

Alignment of human Ubl-family proteins that contain C-terminal Gly-Gly (diGly). Remnant peptides attached to the target lysines after Lys-C digestion (blue) show that mutating the residue before Gly-Gly sequence to lysine (red) as in SUMO2^{T90K} creates an additional cleavage site for Lys-C, which generates a diGly remnant on modified lysines with a unique mass-to-charge (m/z) signature. Lysines targeted by Lys-C are underlined.

digestion of sumoylated proteins will generate a Gly-Gly remnant that can be assumed to be unique to 6His-SUMO2^{T90K}.

Importantly, the T90K variant of SUMO2 can also be exploited to reduce the complexity of the peptide mixture. A monoclonal antibody that specifically recognises Gly-Gly-modified lysines has been developed and can thus be utilised for the enrichment of low abundance 6His-SUMO2^{T90K} remnant-modified peptides after Lys-C digestion (Kim et al., 2011). The purified lower complexity sample is then readily analysed by LC-MSMS. Furthermore, as the Gly-Gly modification (114.0429 Da) is considerably smaller than the remnant of endogenous wild-type SUMO2 (3549.5367 Da), conventional MS software can be efficiently utilised to determine the sequences of modified peptides.

3.2.3. Functional characterisation of SUMO2^{T90K}

To assess the effect of T90K mutation on SUMO2 function, the activities of recombinant SUMO2 and SUMO2^{T90K} were compared in multiple *in vitro* enzyme assays (data and images generated by Ellis G. Jaffray). Four well-established targets of SUMO conjugation: Sp100 (Nuclear autoantigen Sp-100) (Sternsdorf et al., 1997), RanGAP1 (Ran GTPase-activating protein 1) (Matunis et al., 1996; Mahajan et al., 1997), PML (Protein PML) (Sternsdorf et al., 1997), and IRF-2 (Interferon regulatory factor 2) (Han et al., 2008) were purified and subjected to *in vitro* sumoylation analyses with increasing amount of either SUMO2 or SUMO2^{T90K} (Figure 3.2). These assays showed little substrate-dependent variation between SUMO2 and SUMO2^{T90K} in terms of polymerisation or conjugation to target proteins. Separation of the two variants of SUMO2 was also accomplished, provided that the SDS-PAGE was performed using a 10 % Bis-Tris protein gel in a MES buffer solution, which resulted in a slightly slower migration rate of SUMO2^{T90K} (Figure 3.2; control lanes).

Desumoylation assays with 10 nM of SENP1 (Sentrin/SUMO-specific protease SENP1) over a time period of 10 minutes demonstrated that both variants of SUMO2 were deconjugated from the same set of well-established substrate proteins at similar rates (Figure 3.3).

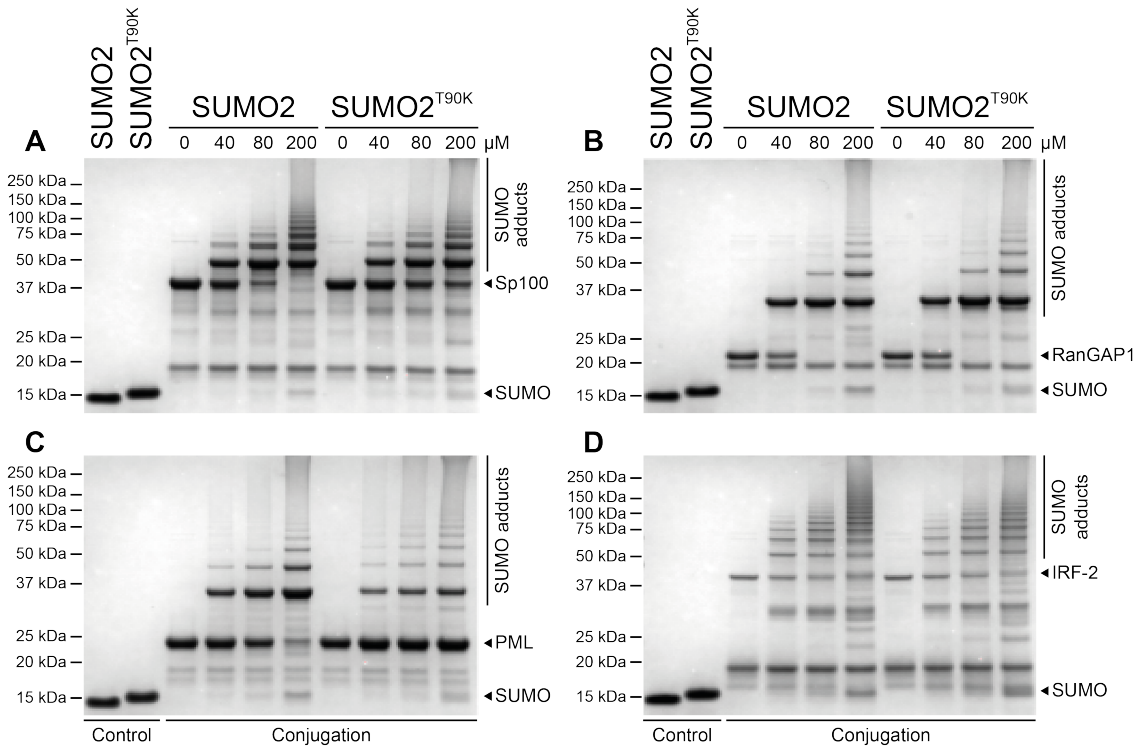

Figure 3.2 | SUMO2 and SUMO2^{T90K} polymerisation and conjugation to various substrate proteins *in vitro* has a similar concentration dependence.

Images of Coomassie brilliant blue-stained SDS-PAGE protein gels comparing the *in vitro* sumoylation of recombinant SUMO2 or SUMO2^{T90K} to four distinct substrate proteins: (**A**) Sp100, (**B**) RanGAP1, (**C**) PML, and (**D**) IRF-2. Courtesy of Ellis G. Jaffray.

Finally, processing of full-length pro-forms of SUMO2 and SUMO2^{T90K} into active mature proteins using 100 nM of SENP1 or 200 nM of SENP2 (Sentrin/SUMO-specific protease SENP2) showed no significant difference irrespective of the protease utilised for the assay (Figure 3.4). These data confirmed the ability of SUMO2^{T90K} to participate in *in vitro* modification, deconjugation and maturation reactions in a comparable manner to the wild-type counterpart.

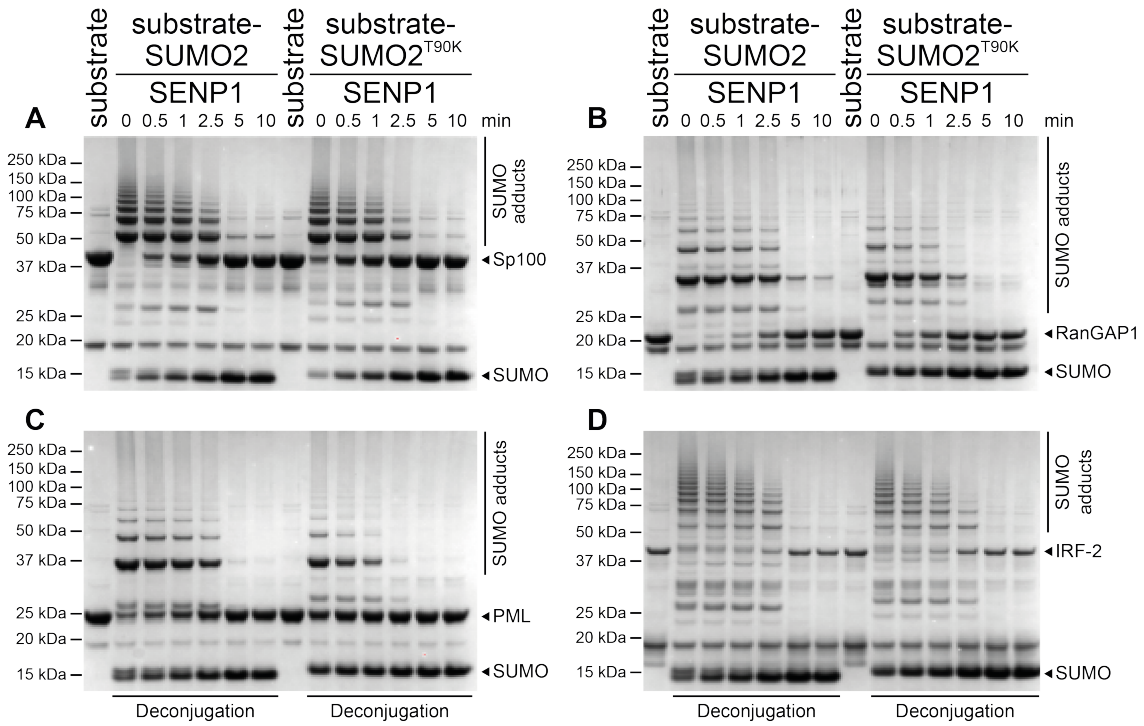

Figure 3.3 | SUMO2 and SUMO2^{T90K} are deconjugated from various substrate proteins at similar rates *in vitro*.

Images of Coomassie brilliant blue-stained SDS-PAGE protein gels comparing the *in vitro* deconjugation of SUMO2 and SUMO2^{T90K} with SENP1 from four distinct sumoylated substrate proteins: (**A**) Sp100, (**B**) RanGAP1, (**C**) PML, and (**D**) IRF-2. Courtesy of Ellis G. Jaffray.

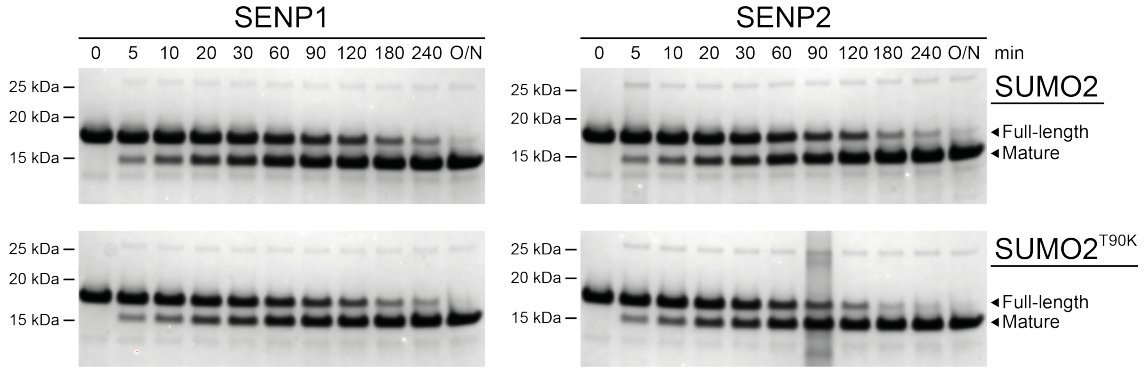

Figure 3.4 | SUMO-specific proteases process the inactive pro-forms of recombinant SUMO2 and SUMO2^{T90K} into functional mature proteins at a similar rate.

Images of Coomassie brilliant blue-stained SDS-PAGE proteins gels comparing the *in vitro* maturation of recombinant SUMO2 (top) or SUMO2^{T90K} (bottom) with SENP1 (left) or SENP2 (right). Courtesy of Ellis G. Jaffray.

3.2.4. Characterisation of HEK293 N3S cells stably expressing 6His-SUMO2^{T90K}

To express SUMO2^{T90K} *in vivo*, human embryonic kidney 293 (HEK293) cells adapted for growth in suspension were used as a parental cell line and stably transfected with a bicistronic eukaryotic expression vector pEFIRESpuro (Hobbs et al., 1998) to generate 6His-SUMO2^{T90K}-expressing HEK293 cells (HEK293$^{6His\text{-}SUMO2\text{-}T90K}$ N3S; implemented by Adel F. M. Ibrahim). Plasmid pEFIRESpuro contains a bicistronic transcription cassette with complementary DNAs encoding the protein of interest (6His-SUMO2^{T90K}) and the selectable marker Puromycin N-acetyltransferase. Transcription will generate a single mRNA molecule with an internal ribosome entry site of an encephalomyocarditis virus downstream of the coding sequence of the desired protein. The expression of 6His-SUMO2^{T90K} is thus ensured in clones that are resistant to the antibiotic puromycin (Hobbs et al., 1998). Using this approach, stable cells expressing 6His-SUMO2^{T90K} were isolated.

Light microscopy was utilised to compare the morphologies of parental and newly generated 6His-SUMO2^{T90K}-expressing HEK293 cells. Photographs of adherent cell cultures with 20 times optical magnification demonstrated that they were indistinguishable (Figure 3.5).

Growth rates were determined by subculturing both of the cell lines in 32 replicates and counting the cells daily over a time period of 2–5 days (Figure 3.6). Irrespective of the utilised culturing conditions, the replicate values of all data points from both cell lines followed the exponential distribution and displayed an expected increase in

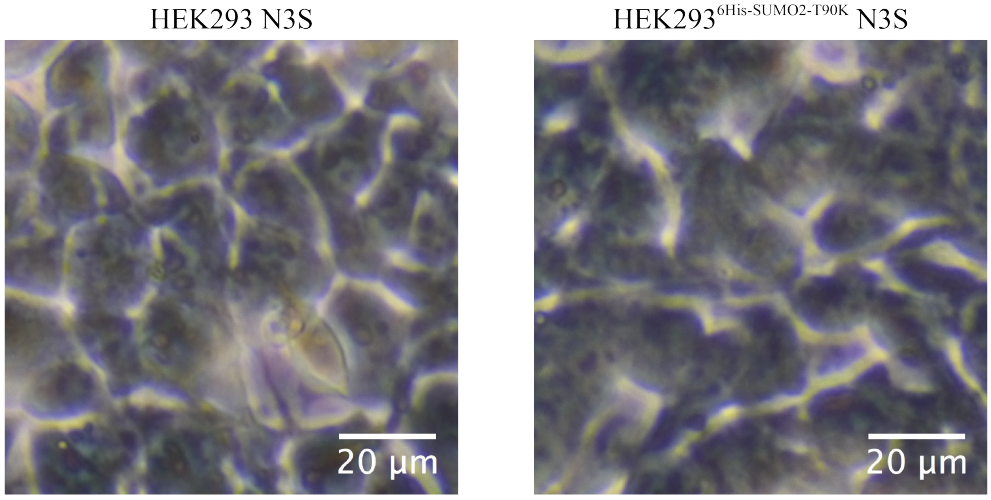

Figure 3.5 | Parental and 6His-SUMO2^{T90K}-expressing HEK293 N3S cells share virtually identical morphology.

Comparison between adherent wild-type HEK293 cells (left) and those stably expressing 6His-SUMO2^{T90K} (right) visualised using light microscopy. Scale bars represent 20 µm.

standard deviation with higher number of cells (Figure 3.6A). However, variation in the initial number of subcultured cells meant that HEK293$^{6His\text{-}SUMO2\text{-}T90K}$ cells grown under puromycin selection reached nearly maximum confluency after 96 hours (Figure 3.6B). The last measurement of this cell culture was thus omitted from the downstream analyses to allow fitting of an exponential model. Best-fit curves describing the exponential growth were then modelled based on the replicate values of individual data points using nonlinear weighted least squares regression. As the variability of replicate data points across the measured time period was heteroscedastic, weighting factor $1/Y^2$ was selected to minimise the sum of the squares of relative distances of replicate values from the curve (Figure 3.6). Doubling times of both cell lines were then calculated from these best-fit curves and appeared to be comparable – 16.1 or 18.5 hours for wild-type or 6His-SUMO2^{T90K}-expressing cells, respectively. However,

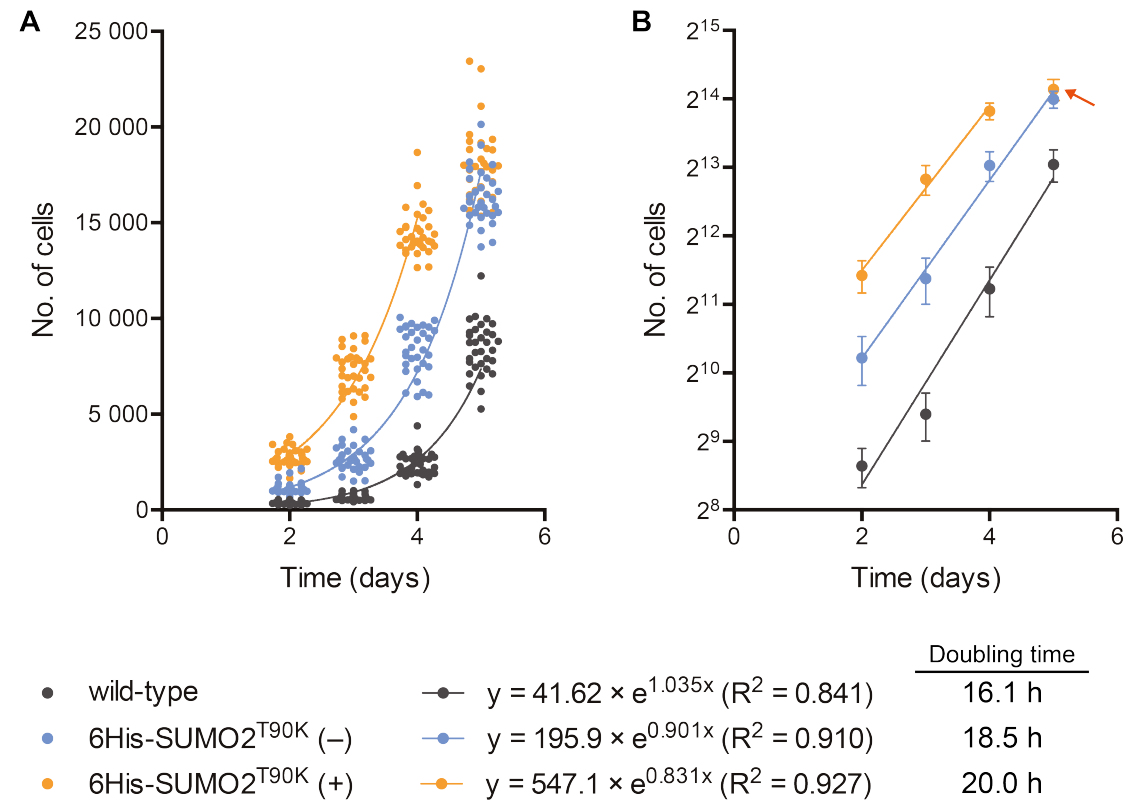

Figure 3.6 | Comparable doubling time can be observed between parental and 6His-SUMO2^{T90K}-expressing HEK293 cells.

The number of cells was counted daily in 32 biological replicates for adherent wild-type HEK293 cells (black) or for those stably expressing 6His-SUMO2^{T90K} cultured either without (–; blue) or under puromycin selection (+; orange) over a time period of 2–5 days after subculturing. (**A**) Scatter plot of individual cell count values. (**B**) Graph of average number of cells with y-axis in logarithmic scale (base 2). All series of data points were fitted with a curve describing exponential growth modelled using nonlinear least squares regression weighted by a factor $1/Y^2$. Error bars correspond to standard deviations and the omitted data point is indicated by a red arrow.

supplementation with the selection antibiotic puromycin (1 µg/ml) had a cumulative effect on the growth rate as indicated by the increase in the doubling time of 6His-SUMO2^{T90K}-expressing cells to 20.0 hours.

Finally, the abundance of 6His-SUMO2^{T90K} was evaluated together with its conjugation in response to cellular stress. Parental and HEK293$^{6His\text{-}SUMO2\text{-}T90K}$ cells were harvested

either unperturbed or after heat stress at 43 °C for 30 minutes. Equivalent amounts of cell lysates (protein wt/wt) were then subjected to Western blot analyses using antibodies recognising either SUMO2/3 or the 6His tag (Figure 3.7). The immunoblot image with SUMO2/3-specific antibody highlighted the near identical amounts of endogenous wild-type and exogenous 6His-SUMO2^{T90K} produced by the HEK293$^{6His\text{-}SUMO2\text{-}T90K}$ cells (Figure 3.7A; lanes 3 and 4). Moreover, enhanced polymerisation and conjugation of 6His-SUMO2^{T90K} to substrate proteins in cells exposed to the elevated temperature was consistent with the response of endogenous SUMO2/3 produced by the parental cells (Figure 3.7). This was confirmed by the Western blot analysis utilising 6His-specific antibody, which enabled to exclusively detect the response of 6His-SUMO2^{T90K} (Figure 3.7B). HEK293$^{6His\text{-}SUMO2\text{-}T90K}$ N3S cells were thus approved as a suitable tool for the identification of SUMO2 modification sites by mass spectrometry.

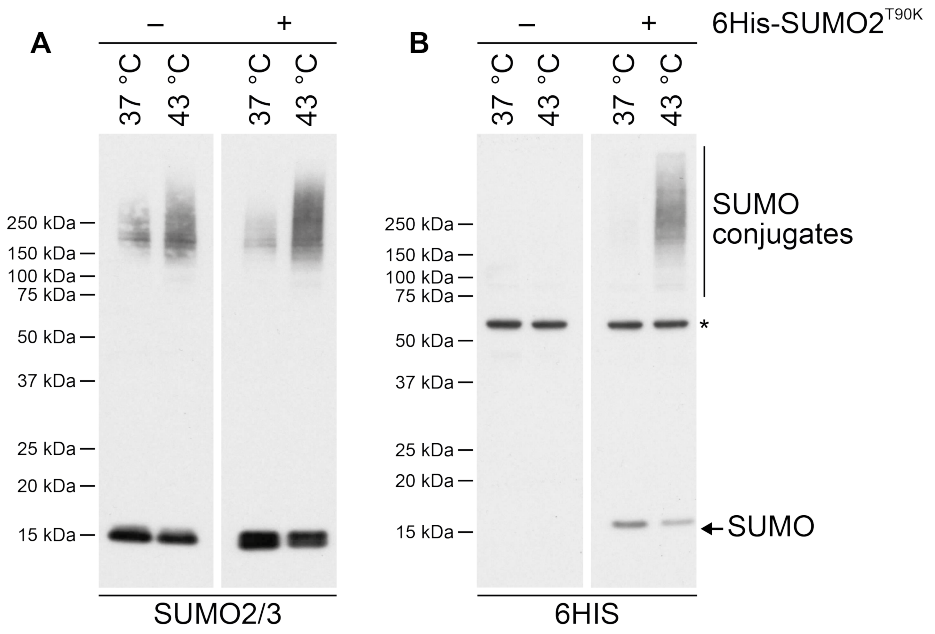

Figure 3.7 | In HEK293$^{6His\text{-}SUMO2\text{-}T90K}$, endogenous SUMO2/3 and 6His-SUMO2^{T90K} are expressed at comparable levels and respond to heat stress similarly.

Lysates from parental (–) or stable 6His-SUMO2^{T90K}-expressing (+) HEK293 cells either unstressed (37 °C) or in response to heat stress (43 °C, 30 min) were subjected to western blot analysis for the detection of: (**A**) SUMO2/3, or (**B**) 6His. Nonspecific immunoreaction is indicated by asterisk.

3.2.5. Overview of the biochemical approach designed for the global identification 6His-SUMO2^{T90K} modification sites

To enable proteome-wide identification of SUMO2 modification sites, a biochemical enrichment approach was developed exploiting the newly characterised HEK293$^{6His\text{-}SUMO2\text{-}T90K}$ cells together with the high sensitivity and throughput of MS-based proteomics. The method (hereafter referred to as the SUMO^{KGG} workflow) employs two tandem purification steps involving the enrichment of sumoylated proteins followed by an affinity purification of peptides modified by the Gly-Gly remnant of 6His-SUMO2^{T90K} (Figure 3.8).

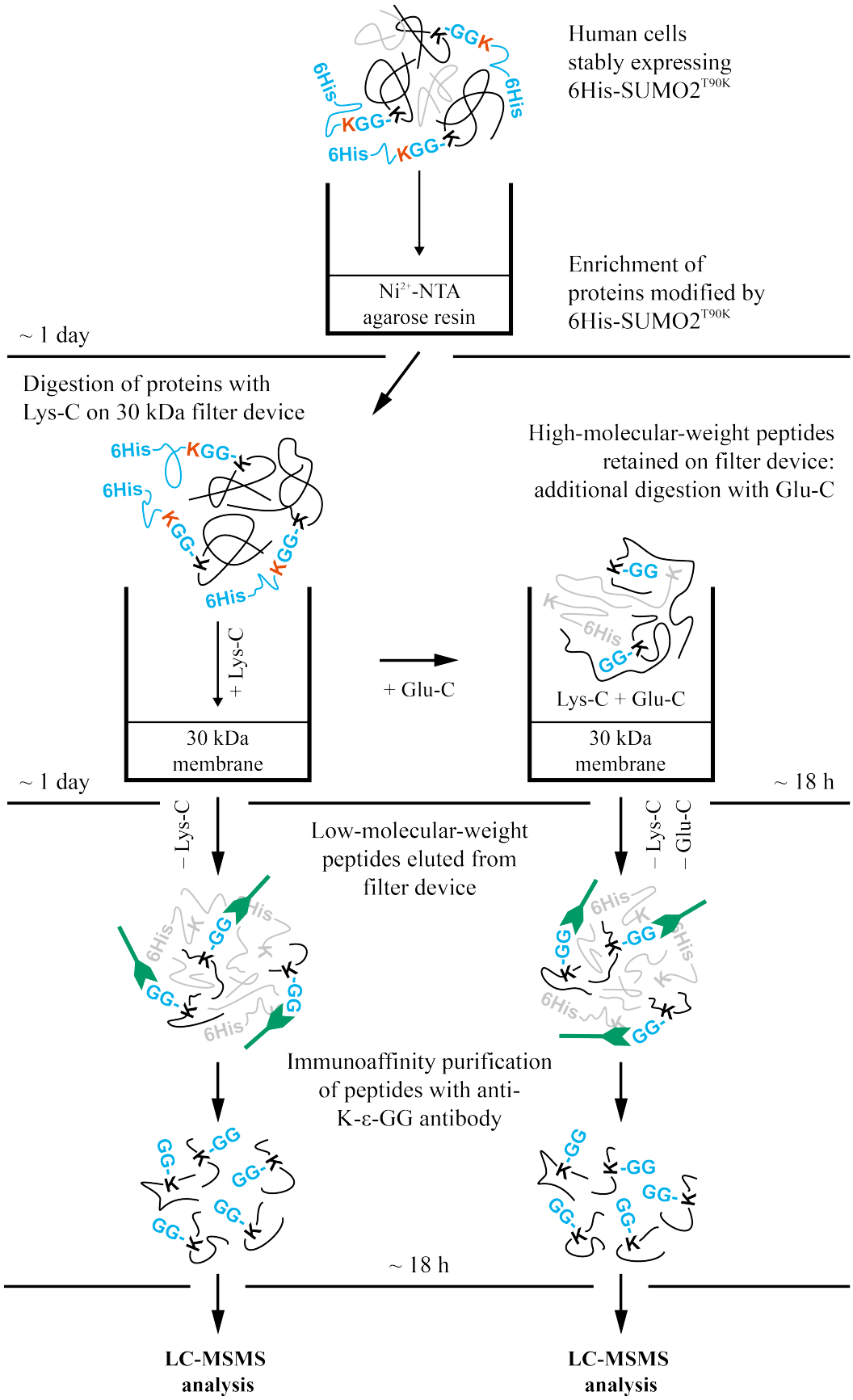

Figure 3.8 | Depiction of the SUMO^{KGG} workflow.

››

‹‹ **Figure 3.8 | Depiction of the SUMO^{KGG} workflow.**

Proteins conjugated to 6His-SUMO2^{T90K} are purified using nickel affinity chromatography, digested with Lys-C and resulting peptides are separated from the protease using a 30 kDa filter device. High-molecular-weight peptides retained on the device are subsequently cleaved with Glu-C, filtered through the membrane and samples are kept separately. Finally, SUMO remnant-containing peptides are enriched from each sample using a K-ε-GG-specific antibody and analysed by LC-MSMS.

In this strategy, proteins are extracted by cell lysis under strong denaturing conditions to inactivate SUMO proteases and prevent desumoylation of substrates. Cell lysate proteins are then mixed with Ni^{2+}-NTA agarose resin and stirred overnight at 4 °C to allow specific binding of the polyhistidine stretch of 6His-SUMO2^{T90K} to the divalent nickel ion (Ni^{2+}), immobilised to an agarose matrix via a coordination complex with nitriloacetic acid (NTA) (Hochuli et al., 1987; Bornhorst and Falke, 2000). Next, the slurry of Ni^{2+}-NTA agarose with cell lysate proteins is transferred into an empty spin chromatography column and agarose beads are allowed to settle to the bottom of the container by gravity. Packed Ni^{2+}-NTA resin is then stringently washed while altering the pH of the buffer solutions for the elution of nonspecifically bound proteins. Importantly, the pH of these buffers should not be adjusted below 6.0, as the imidazole ring of the histidine residue becomes protonated. This phenomenon disrupts the coordination complex and results in a premature dissociation of polyhistidine-tagged proteins from the resin (Bornhorst and Falke, 2000). Finally, 6His-SUMO2^{T90K} conjugates are separated from the Ni^{2+}-NTA stationary phase using an imidazole-based eluent that competes with the imidazole rings of the histidines and enables the efficient release of specifically bound substrate proteins (Figure 3.8; top panel).

Proteins eluted from the Ni^{2+}-NTA agarose resin are then concentrated in an ultrafiltration device (Figure 3.8; second panel). This device acts as a container while subsequent sample preparation procedures are applied, including proteolysis into peptides. The nominal molecular weight cutoff limit should thus be carefully selected to obtain the desired properties of separation. An appropriate cutoff assists a rapid buffer exchange for the complete removal of chemicals incompatible with the subsequent stages of the protocol, and ultimately facilitates the efficient elution of derivative peptides. On the other hand, partially or undigested proteins, and high-molecular-weight peptides should be retained on the filter device, as they can interfere the eventual analysis by LC-MSMS (Manza et al., 2005; Wisniewski et al., 2009b). Most importantly, utilised proteases should be efficiently separated from the peptide mixture to prevent cleavage of diGly-Lys (K-ε-GG)-specific antibody. Thus, a 30 kDa ultrafiltration device was employed and its suitability was comprehensively examined in Chapter 3.2.7.

Prior to proteolytic digestion, the concentrated denatured proteins in the ultrafiltration device are treated with 2-chloroacetamide to carbamidomethylate the reduced thiol groups of cysteine residues and block the subsequent formation of disulfide bonds (Figure 3.8; second panel). The advantage of 2-chloroacetamide over the routinely utilised 2-iodoacetamide is its high substrate selectivity, which prevents the derivatisation of lysine residues with 2-acetamidoacetamide adducts that have an identical atomic composition to Gly-Gly remnant ($C_4H_6N_2O_2$; 114.0429 Da) and are thus not distinguished by MS (Nielsen et al., 2008). The excess of 2-chloroacetamide is removed by applying stringent washes and eventually replaced with immunoaffinity purification (IAP) buffer. Proteolysis in IAP buffer streamlines the sample handling

and ultimately improves the yield of peptides, as the arising mixture can be directly subjected to the second enrichment procedure.

Lys-C is essential to the SUMO^{KGG} workflow owing to its high specificity towards the hydrolysis of the peptide bonds at the C-terminal side of lysines alone (Raijmakers et al., 2010). This will ensure that substrates conjugated to 6His-SUMO2^{T90K} will be left with the Gly-Gly remnant attached to the target lysines. As other human Ubls do not contain a lysine in the equivalent position (Figure 3.1), diGly-modified peptides can only originate from 6His-SUMO2^{T90K}. Finally, these Lys-C-digested peptides in IAP buffer are eluted from the ultrafiltration device (Figure 3.8; second panel).

Covalent modification of lysine ε-amino groups renders the sites resistant to Lys-C digestion (Giansanti et al., 2016). It was thus expected that lysine-deficient regions of proteins could generate peptides too large to pass through the membrane of the ultrafiltration device or for reliable identification by LC-MSMS. To increase the proteome coverage, a second endoproteinase Glu-C was introduced into the workflow (Figure 3.8; second panel), which cleaves at the C-terminal side of glutamic and aspartic acids (Drapeau et al., 1972). Glu-C will not affect the Gly-Gly remnant of 6His-SUMO2^{T90K}, but can shorten the high-molecular-weight peptides retained on the membrane. These truncated peptides generated by the subsequent Glu-C digestion are also eluted from the filter device, but stored separately from the initial subfraction of Lys-C cleaved peptides (Figure 3.8; second panel).

Both subfractions of peptides are then independently subjected to the second enrichment procedure using a commercially available monoclonal antibody specific for

Gly-Gly-modified lysines (Kim et al., 2011). Selective separation of diGly-modified species from the excess of unmodified peptides is required to reduce the complexity of the samples (Figure 3.8; third panel). Finally, these purified Gly-Gly-Lys-containing peptide sequences are identified and the modification sites are determined using LC-MSMS (Figure 3.8; bottom panel).

3.2.6. Purification of proteins conjugated to 6His-SUMO2^{T90K}

Three members of the human Ubl family contain an arginine at the equivalent position to the T90K mutation in 6His-SUMO2^{T90K} (Figure 3.1). Considering the low stoichiometry of sumoylated proteins, it was speculated that their separation from the excess of proteins conjugated to any of the residual Ubls could be beneficial. In particular, if unspecific digestion of Lys-C at the C-terminal side of arginine residues occurs even with low efficiency, a considerable number of identified Gly-Gly modification sites could in reality emerge from Ubiquitin, ISG15 or NEDD8, instead of 6His-SUMO2^{T90K}.

Secondly, the K-ε-GG-specific monoclonal antibody that is utilised in the subsequent stage of the SUMO^{KGG} workflow, was originally designed for the enrichment diGly-Lys-containing peptides of trypsin-digested ubiquitinated proteins (Kim et al., 2011). Numerous valuable characteristics of the antibody, such as specificity or binding capacity have thus been determined, and the established protocols can be adjusted for the enrichment of SUMO remnant-modified peptides. For instance, merely 31 µg of K-ε-GG-specific antibody conjugated to 5 µl of protein A beads is required for the enrichment of ubiquitinated peptides from 10 mg of MG132-treated cell lysate proteins. Interestingly, utilisation of higher amount of antibody has a negative effect, as it

increases the proportion of nonspecifically bound unmodified peptides, which then hamper the identification of genuine targets (Udeshi et al., 2013b). In human cells, Ubiquitin is present in approximately 20-fold excess compared to the absolute abundance of SUMO paralogs (Kaiser et al., 2011; Matic et al., 2011; Wisniewski et al., 2014). It was thus predicted that less than 31 μg of antibody is required for the optimal enrichment of SUMO2^{T90K} remnant-modified peptides. Therefore, the lowest amount of antibody-conjugated resin that could be reproducibly handled was established and corresponded to 3 μl of protein A beads conjugated to ~19 μg of K-ε-GG-specific antibody. Such a low quantity of resin, however, generated a demand to work with low-volume peptide samples to prevent the loss of antibody-bound beads to the sides of a large container. The efficient enrichment of SUMO2^{T90K} conjugates could simultaneously decrease the complexity and reduce the volume of protein samples.

To evaluate the suitability of nickel affinity chromatography for the enrichment of 6His-SUMO2^{T90K} conjugates, proteins extracted from heat-stressed HEK293$^{6His\text{-}SUMO2\text{-}T90K}$ cells were combined with Ni^{2+}-NTA resin in a ratio of 200:1 (wt/vol). After an overnight rotation at 4 °C, the slurry was transferred into an empty spin chromatography column and the residual solution of unbound proteins was collected by gravity. Nonspecifically bound proteins were removed by repeatedly washing the column with buffer solutions at different pH, and 6His-SUMO2^{T90K}-bound proteins were eluted in three sequencial steps (a detailed protocol is provided in Chapter 2.3.1). An equivalent proportion of the volume of each fraction was then loaded onto an SDS-PAGE protein gel and subjected to Western blot analyses for the detection of SUMO2/3 or 6His (Figure 3.9).

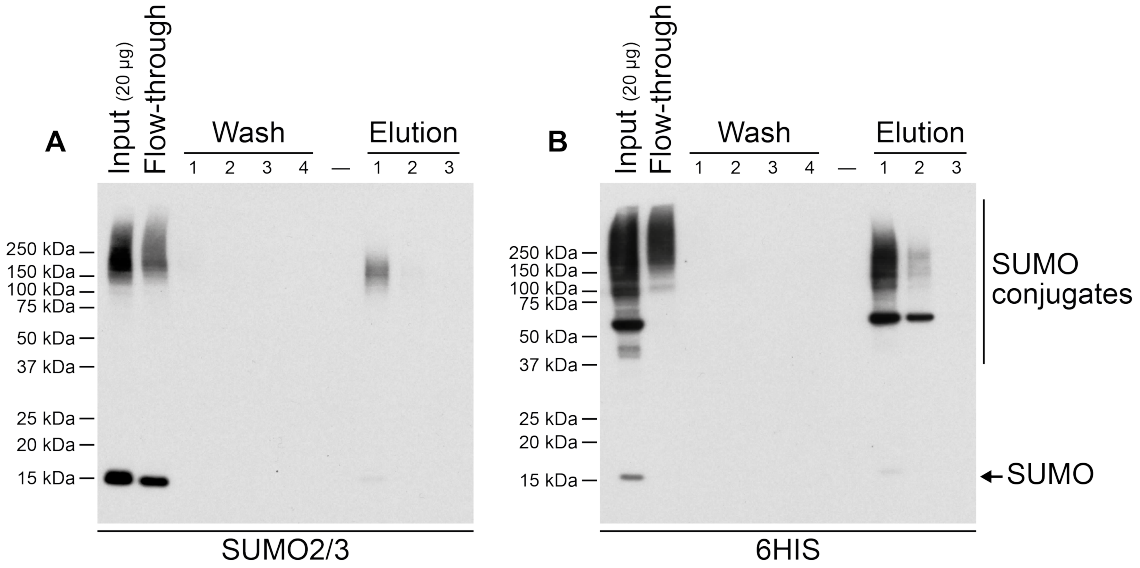

Figure 3.9 | 6His-SUMO2^{T90K}-conjugated proteins can be enriched using nickel affinity chromatography.

Equal proportion (~0.04 %; vol/vol) of HEK293$^{6His\text{-}SUMO2\text{-}T90K}$ cell lysate (input) and each nickel affinity chromatography fraction (flow-through; washes 1–4; elutions 1–3) were subjected to Western blot analysis for the detection of: (**A**) SUMO2/3 or (**B**) 6His. Approximately 0.04 % of input material corresponds to 20 µg of protein. Empty lane (_).

A SUMO2/3-specific immunoblot image illustrated the reduced abundance of sumoylated proteins in the high-molecular-weight area of the flow-through fraction of the nickel affinity chromatography. Likewise, the amount unconjugated SUMO2/3 was also decreased (Figure 3.9A). Notably, HEK293$^{6His\text{-}SUMO2\text{-}T90K}$ cells express endogenous SUMO2/3 and 6His-SUMO2^{T90K}, which compete for conjugation. Therefore, the residual sumoylated proteins in the flow-through fraction were likely to be the targets of endogenous SUMO2/3. Finally, the immunoblot image of SUMO2/3 confirmed that modified proteins were not released from the Ni^{2+}-NTA agarose matrix during the stringent wash procedure, but could be efficiently eluted with imidazole (Figure 3.9A).

6His-specific Western blot analysis illustrated that the substrates of 6His-SUMO2^{T90K} were found in elutions 1 and 2 (Figure 3.9B). Consistent with the SUMO2/3-specific

immunoblot, the amount of sumoylated proteins was decreased and unconjugated 6His-SUMO2^{T90K} was undetectable in the flow-through fraction. However, a considerable amount of antibody-reactive proteins were still detected among the unbound proteins (Figure 3.9B; lane 2). Therefore, another experiment was performed to investigate whether these were substrates, which could be recovered by conducting repeated Ni^{2+}-NTA purifications. This set-up mimicked the conditions created by adding twice as much resin to the cell lysate protein, but diminished the probability of rendering the samples unsuitable for further investigation by co-purifying extensive quantities of nonspecifically-bound contaminants.

This time, HEK293$^{6His\text{-}SUMO2\text{-}T90K}$ cells were treated with an agent GRE010, which is a suspected inhibitor of Ubl proteases, including SENPs. As predicted, the amount of SUMO2-bound substrates in the high-molecular-weight area of the protein gels was increased in response to the treatment with 20 μM GRE010 for 8 hours, whereas the cells mock-treated with 0.2 % dimethyl sulfoxide (DMSO) were not affected (Figure 3.10; lanes 1 and 2). Similarly to the nickel affinity chromatography with the lysate protein of heat-stressed cells, a considerable amount of residual SUMO substrates were detected from the flow-through fraction of the first nickel affinity chromatography procedure (Figure 3.10; flow-through α). However, submission of this fraction of unbound proteins into second purification cycle did not lead to a considerable amount of additionally enriched substrates of 6His-SUMO2^{T90K} (Figure 3.10; elution β, 1–3).

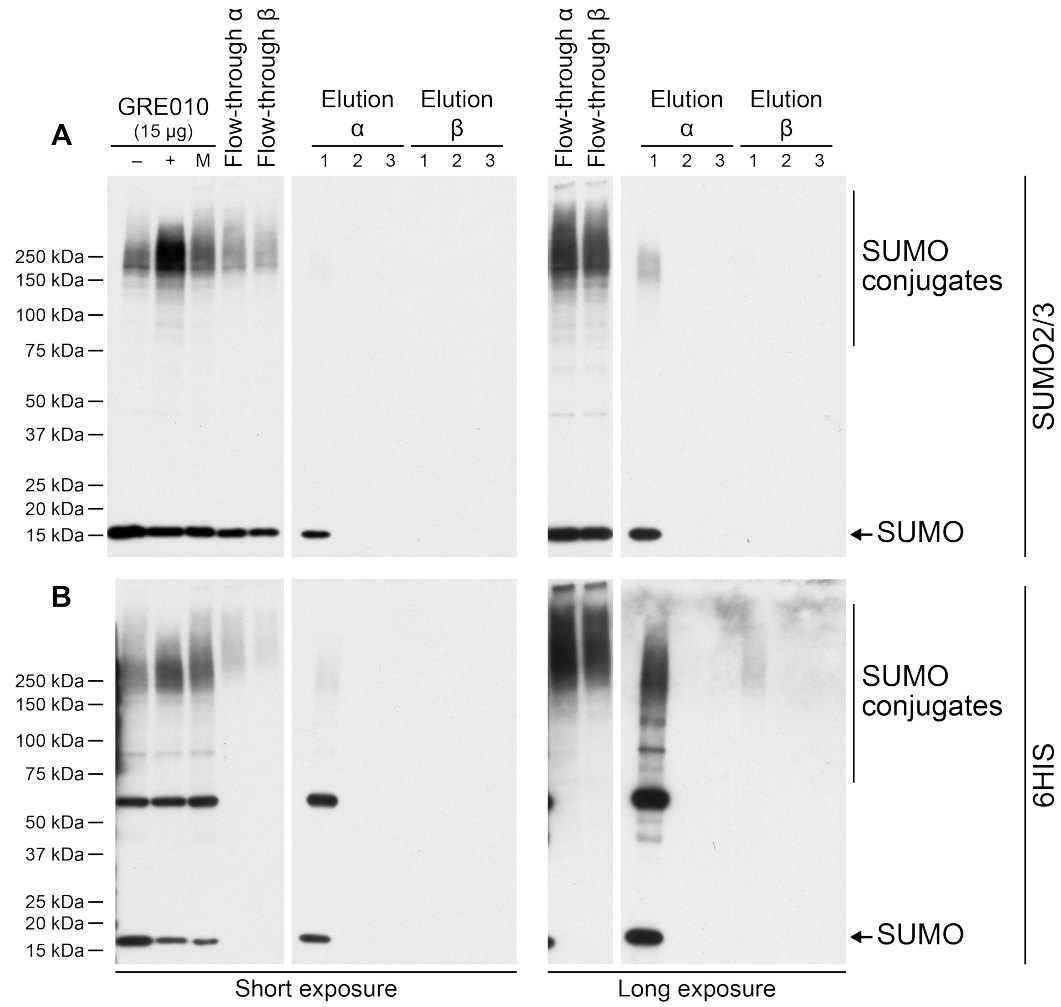

Figure 3.10 | Sumoylated or polyhistidine-containing proteins in the flow-through fraction are not efficiently retained during the second cycle of nickel affinity chromatography.
HEK293$^{6His\text{-}SUMO2\text{-}T90K}$ cells were either incubated with 20 µM GRE010 for 8 hours (GRE010; +) or mock-treated with 0.2 % DMSO (GRE010; –). Both cell cultures were harvested and equal amount of each cell pellet (wt/wt) was mixed (GRE010; M). 6His-SUMO2^{T90K}-conjugated proteins were extracted from the combined sample utilising two Ni^{2+}-NTA chromatography procedures in tandem (α and β). Equivalent proportion (~0.02 %; vol/vol) of each fraction was subjected to western blot analysis for the detection of: (**A**) SUMO2/3 or (**B**) 6His. Approximately 0.02 % of each cell lysate (GRE010) corresponds to 15 µg of protein.

Overall, comparisons between the cell lysate and nickel affinity chromatography elution fraction revealed that the enrichment of 6His-SUMO2^{T90K} conjugates using a single procedure resulted in ~20-fold and ~130-fold reduction in sample volume or the amount of protein, respectively.

3.2.7. Depletion of endoproteinases Lys-C and Glu-C from the solution of peptides

As the 6His-SUMO2^{T90K}-modified proteins are cleaved prior to the second enrichment of sumoylated peptides, it is crucial to separate the samples from the proteases to avoid digesting the K-ε-GG-specific antibody. Filter-aided sample preparation (FASP) protocols employing ultrafiltration devices with specified nominal molecular weight cutoff limits have been widely used to separate proteins and peptides from low-molecular-weight compounds or high-molecular-weight substances that can interfere with MS analyses (Manza et al., 2005; Wisniewski et al., 2009b). The molecular weights of endoproteinases Lys-C and Glu-C are ~28 kDa or ~30 kDa, respectively (The UniProt Consortium, 2015). It was thus anticipated that both of these proteases might be retained on an ultrafiltration device with a nominal molecular weight cutoff limit of 30 kDa or smaller. However, previous comparative studies have shown that ultrafiltration devices with larger nominal cutoff limits (30 kDa or 50 kDa) produce peptides with higher yield (Wisniewski et al., 2011). For these reasons, a 30 kDa ultrafiltration device was selected and comprehensively investigated for the compatibility with the SUMO^{KGG} workflow.

The retention of Lys-C and Glu-C was examined by placing 15 µg of bovine serum albumin (BSA) tryptic peptides either alone or combined with 5 µg of Lys-C (Figure 3.11A) or 5 µg of Glu-C (Figure 3.11B) onto a 30 kDa ultrafiltration device in a final volume of 50 µl of IAP buffer. Each of these mixtures was then processed and passed through the filter using identical experimental conditions to the SUMO^{KGG} workflow, and flow-through fractions were collected. Each of the filter-retained fraction was then

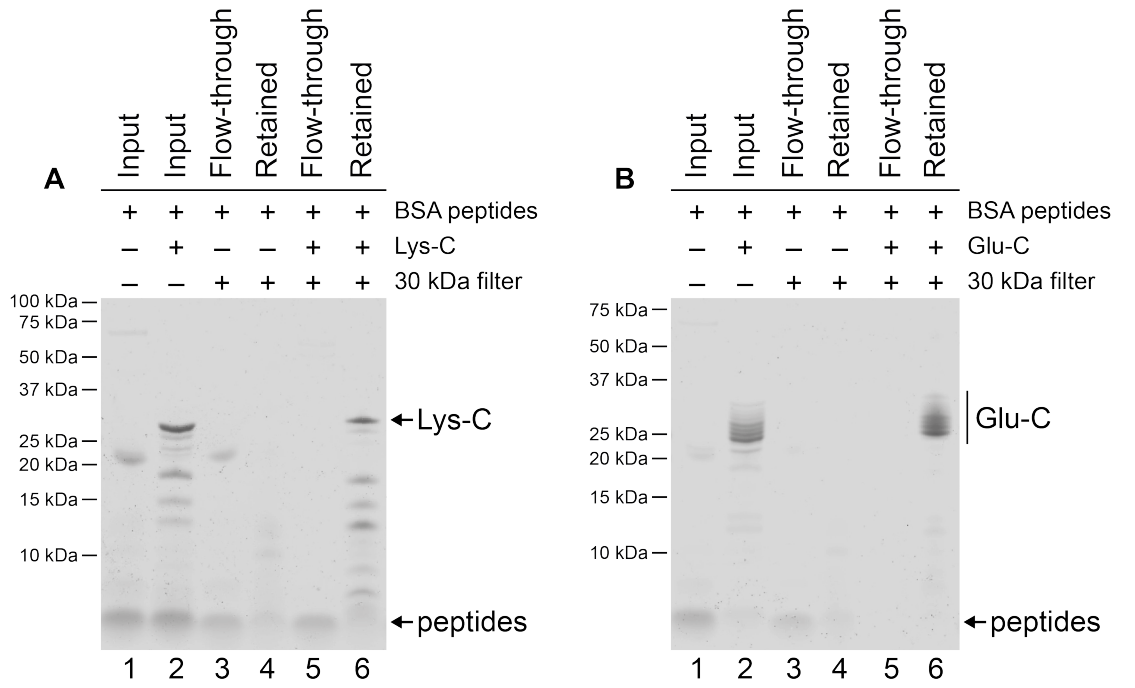

Figure 3.11 | Endoproteinases Lys-C and Glu-C are sufficiently separated from a mixture of BSA peptides using 30 kDa filter units.

A solution of BSA tryptic peptides either alone, or mixed with Lys-C (**A**) or Glu-C (**B**) was passed through 30 kDa cutoff filter unit. Proteins and peptides retained on the filter (lanes 4 and 6), collected from the flow-through fraction (lanes 3 and 5), or without fractionation (Input; lanes 1 and 2) were analysed using Coomassie brilliant blue-stained SDS-PAGE protein gels.

retrieved by the addition of IAP buffer and reverse centrifugation of the ultrafiltration device. As a control, a second collection of mixtures was produced by dissolving the individual components directly into IAP buffer, which were then examined without passing through the 30 kDa ultrafiltration device. Finally, all generated samples were concentrated, dissolved in 1 × Laemmli sample buffer (125 mM Tris-HCl pH 6.7, 2 % SDS, 15 % glycerol, 300 mM DTT) and analysed by SDS-PAGE (Figure 3.11).

Images of Coomassie brilliant blue-stained SDS-PAGE protein gels indicated that both proteases, Lys-C and Glu-C, were retained on the 30 kDa filter device, while tryptic BSA peptides were predominantly detected from the flow-through fractions (Figure

3.11). Upon the supplementation of Glu-C, BSA peptides present in the sample were truncated and thus remained unresolved on the SDS-PAGE protein gel (Figure 3.11B; lanes 3 and 5). Interestingly, Coomassie brilliant blue-stainable material with significantly smaller molecular weight than the 30 kDa cutoff limit was detected from "input" and "retained" fractions, which also contained either Lys-C or Glu-C (Figure 3.11; lane 6). As BSA was successfully separated from the Lys-C containing sample (Figure 3.11A; lane 5), it is likely that the fragments emerged from the autolytic digestion of the protease during the relatively slow solvent evaporation procedure immediately before SDS-PAGE analysis. However, alternative hypotheses, such as the possibility that these low-molecular-weight substances were selectively retained on the 30 kDa ultrafiltration device, cannot be rejected. Nevertheless, a 30 kDa ultrafiltration device can be utilised for the removal full-length proportion of proteolytically active Lys-C and Glu-C from the peptide mixtures.

3.2.8. Characterisation of Lys-C and Glu-C substrate specificity and proteolytic efficiency in IAP buffer

As illustrated in Chapter 3.2.7, utilisation of 30 kDa ultrafiltration device is required for the depletion of Lys-C and Glu-C from peptide mixtures to prevent the digestion of subsequently supplemented K-ε-GG-specific antibody. In principle, this ultrafiltration device acts as a storage chamber for the proteins, while the wash solutions can be passed through the filter. Likewise, this container could be utilised to exchange the protein solution to IAP buffer. Proteolysis in IAP buffer would integrate sample handling and thus increase the recovery of peptides, as the arising solution could be directly subjected to the subsequent enrichment procedure.

Previous studies have shown that both, Lys-C and Glu-C are highly specific endoproteinases that hydrolyse peptide bonds from the C-terminal side of lysines, or glutamic and aspartic acids, respectively (Drapeau et al., 1972; Raijmakers et al., 2010). Nevertheless, the performance of these proteases can be affected by the components of the IAP buffer. Notably, the apparent preferential selectivity of Glu-C towards Glu or Asp residues appears to depend on the concentration of ammonium bicarbonate or ammonium acetate that have inhibitory effects on the activity of the protease (Houmard and Drapeau, 1972; Sorensen et al., 1991). To characterise the enzymatic activities and substrate specificities of both proteases in IAP buffer, HEK293 N3S cell lysate proteins were processed using the developed workflow. Digestion was performed on a 30 kDa filter device using Lys-C, either alone or in tandem with Glu-C in two alternative buffer conditions: IAP buffer, or widely-utilised and well-characterised ammonium bicarbonate (NH_4HCO_3). Peptides from all four conditions were then individually analysed by LC-MSMS.

The specificity of proteases in the described experimental set-up was determined by evaluating the occurrences of individual amino acids at the C-termini of identified peptide sequences (Figure 3.12A). To avoid bias in the identified sequences, all MSMS peak lists were searched using unspecific enzyme settings with the Andromeda search engine integrated into MaxQuant software package (Cox and Mann, 2008; Cox et al., 2011). Approximately 91.5 % of the peptides identified from the sample digested with Lys-C alone in NH_4HCO_3 contained a lysine at the C-terminus (Figure 3.12A; light blue). Interestingly, this proportion increased to 94 % when the digestion was performed in IAP buffer. Moreover, less than 1.5 % of the Lys-C digested peptides

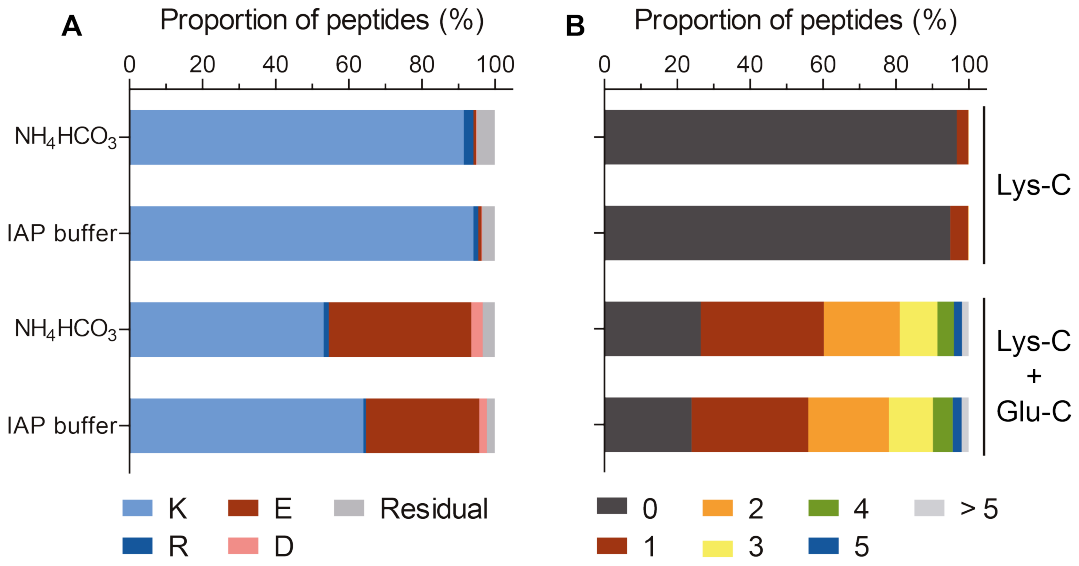

Figure 3.12 | Digestion of proteins in immunoaffinity purification buffer does not alter the efficiency or specificity of endoproteinases Lys-C and Glu-C.

HEK293 cell lysate proteins were prepared according to the FASP protocol using 30 kDa cutoff filter units and protein buffer solution was exchanged either to immunoaffinity purification (IAP) buffer or to 25 mM ammonium bicarbonate (NH_4HCO_3). Proteins were digested either with Lys-C only, or using Lys-C and Glu-C in tandem. Corresponding peptides were analysed by LC-MSMS and raw MS files were searched with MaxQuant software package using nonspecific enzyme settings. (**A**) Stacked bar graph showing the proportion of identified peptide sequences in each of the samples terminating with lysine (K; light blue), arginine (R; dark blue), glutamic acid (E; red), aspartic acid (D; pink), or any other residue (Residual; grey). (**B**) Stacked bar graph demonstrating the proportion of identified peptide sequences in each of the protein digests containing either 0 (dark grey), 1 (red), 2 (orange), 3 (yellow), 4 (green), 5 (blue) or more than five (light grey) sites of missed enzyme cleavages.

identified from IAP buffered sample terminated with an arginine (Figure 3.12A; dark blue) and the proportion of sequences finishing with any other amino acid was less than 5 %. These results confirmed that Lys-C digestion could be performed in IAP buffer, as the enzyme remained highly specific towards the peptide bonds C-terminal to lysine residues.

Similarly, the combined specificity of Lys-C and Glu-C enzymes was slightly improved in IAP buffer with ~95.1 % of the identified peptide sequences terminating with lysine or glutamic acid (Figure 3.12A; light blue and red). This was 3 % greater than the proportion of Lys and Glu at the C-termini of peptides cleaved in ammonium bicarbonate. Interestingly, irrespective of the components of the digestion buffer, an average of ~2.6 % (± 0.5 %) of the sequences contained an Asp at their C-terminus (Figure 3.12A; pink). Thus, the C-termini of most double-cleaved peptides appeared to be occupied either by Lys or Glu, or to a smaller extent by Asp (Figure 3.12A).

Next, the activity of each enzyme in IAP buffer was evaluated by counting the occurrences of missed cleavages in the sequences of identified peptides, while considering the established enzyme specificities (Figure 3.12B). Irrespective of the buffer environment, more than 94.8 % of the sequences cleaved with Lys-C alone contained no internal lysine residues. An average of 4 % of the peptides carried one missed cleavage site and this proportion was only 0.9 % higher in samples digested in IAP buffer (Figure 3.12B). However, ~74.8 % (± 1.3 %) of peptides emerging from the samples digested with Lys-C and Glu-C in tandem contained one or more missed cleavage sites, whereas only a quarter (~25.2 % ± 1.3 %) of the identified peptides were digested completely. Furthermore, the proportion of sequences with up to four missed cleavages was ~95.8 % (± 0.2 %). However, similarly to the Lys-C digested sample, the effect of the buffer solution on the activities of both enzymes when utilised in tandem, was negligible (Figure 3.12B).

In general, peptide sequences are identified by comparing their masses and MSMS fragmentation spectra against a database of in silico-generated theoretical peptide

spectra (Sadygov et al., 2004; Steen and Mann, 2004). Therefore, the identification of any peptide sequence relies on the presence of a corresponding theoretical spectrum in the computer-modelled database. This database should thus be constructed while taking into account the activities and specificities of utilised proteases, to ensure the maximal resemblance between the collections of theoretical and *in vivo*-produced peptides. Characterisation of both enzymes, Lys-C and Glu-C, demonstrated their incomplete activity (Figure 3.12B). In these situations, allowing up to two missed cleavages for the creation of in silico database is widely accepted (Giansanti et al., 2016). However, increasing the number of missed cleavages even further is generally not recommended, as it will generate a greater number of theoretical peptide masses and thus increase the probability of observing a peptide-to-spectrum match (PSM) at random (Eidhammer et al., 2007). To this end, up to two missed cleavages were allowed for the analysis of peptides digested with Lys-C alone. On the other hand, utilisation of these search settings would identify only 79.6 % (± 1.5 %) of the peptides previously discovered from the double-digested test sample (Figure 3.12B). Despite increasing the search space and the probability of random matching, it was considered that allowing up to four missed cleavages would increase the proportion of expected identifications to at least 95 %. Interestingly, these observations are consistent with another independent study (Giansanti et al., 2016).

Finally, the inability of Lys-C to cleave C-terminal to methylated or acetylated lysines has been observed (Giansanti et al., 2016). Sumoylation is another lysine-specific modification and a covalent attachment to SUMO is likely to render the modified lysines resistant to Lys-C cleavage. However, to avoid increasing the search space excessively, the number of allowed missed cleavages was increased by one. All

subsequent analyses of raw MS files were thus performed by allowing up to three or five missed cleavages for samples digested with Lys-C alone, or in combination with Glu-C, respectively.

3.2.9. Evaluating the impact of tandem protein digestion on the sequence coverage of the proteome

To measure whether a sufficient proportion of the proteome is covered by the identified peptide sequences, acquired MS files were re-analysed with MaxQuant using enzyme specificity settings as established in Chapter 3.2.8. Additionally, a sample of peptides produced by trypsin digestion in NH_4HCO_3 was included as a control to determine the proportion of the proteome identified when samples were processed according to a well-established approach (Wisniewski et al., 2009b).

On average, ~57 500 MSMS scans were acquired per peptide sample by the Q Exactive mass spectrometer (Figure 3.13A), and this value differed by less than 2500 events between the individual LC-MSMS runs. However, despite the comparable number of acquired MSMS scans, a considerable protease-dependent discrepancy in the amount of peptide-to-spectrum matches (PSMs) was observed (Figure 3.13A; white). As expected, the digestion with trypsin led to a maximum number of PSMs (~35 000). Irrespective of the composition of the digestion buffer, the number of matched sequences from samples cleaved with Lys-C alone was only 4000 less (~31 000). When Lys-C digested peptides were subsequently treated with Glu-C in NH_4HCO_3 the number of PSMs decreased by ~20 % (to ~25 000). However, this substantial decrease in PSMs was not observed when the digestion was carried out in IAP buffer, as almost 30 000 PSMs were obtained (Figure 3.13A).

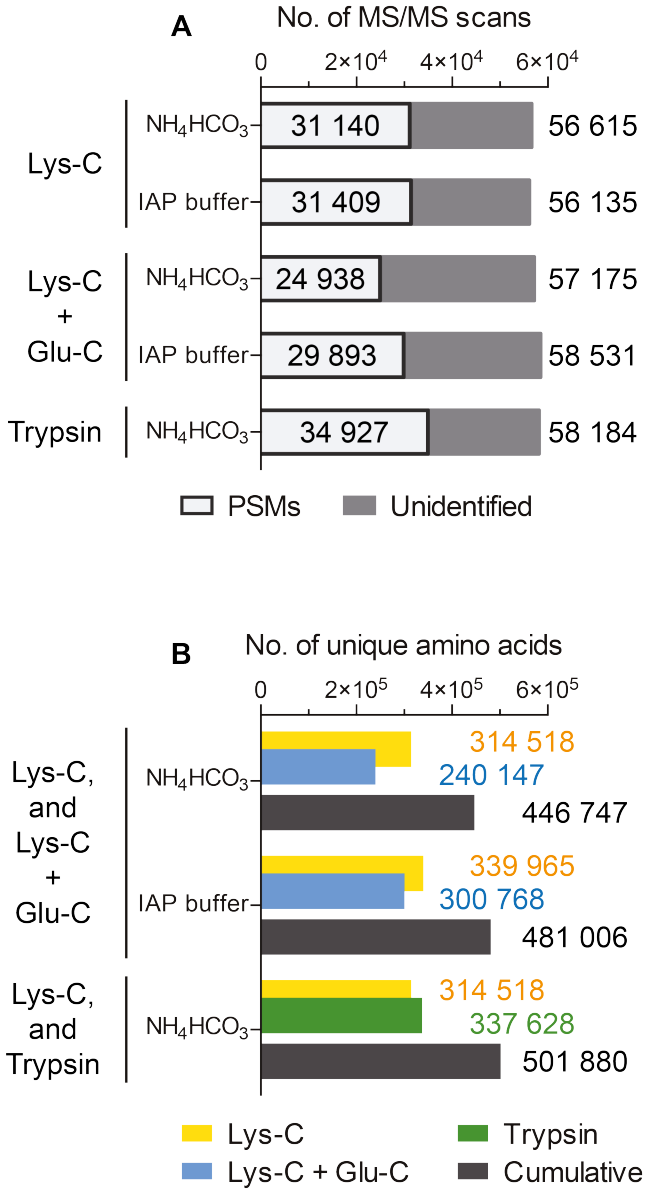

Figure 3.13 | In-parallel digestion of proteins with Lys-C, and with Lys-C and Glu-C contributes to the cumulative increase in proteome sequence coverage.

HEK293 cell lysate proteins with buffer solutions exchanged either to immunoaffinity purification (IAP) buffer or to 25 mM ammonium bicarbonate (NH_4HCO_3) were digested with Lys-C, Lys-C and Glu-C, or trypsin according to the FASP protocol (Manza et al., 2005; Wisniewski et al., 2009b). Corresponding peptides were analysed by LC-MSMS and searched using MaxQuant. (**A**) Stacked bar graph illustrating the number of peptide-to-spectrum matches (PSMs; white) and unidentified MSMS scans (grey) from each of the peptide mixtures. (**B**) Bar graph demonstrating the number of identified nonredundant amino acids covering the proteome sequence in each of the proteolytic digests (Lys-C, blue; Lys-C and Glu-C, red; trypsin, green) or after multiple data sets are combined (Cumulative; grey).

The sequence coverage of each proteome originating from five alternative combinations of proteolytic digests was determined by calculating the proportions of nonredundantly identified amino acid residues (Figure 3.13B). In general, a positive correlation between these results and the number of PSMs was detected. The greatest proportion of the proteome was covered by peptides identified from the trypsin-digested sample, which contained ~340 000 unique amino acid residues (Figure 3.13B; green). Whilst this value was marginally lower when proteins were digested in NH_4HCO_3 with Lys-C alone (~310 000), essentially identical results to trypsin were obtained by proteolysis in IAP buffer, as illustrated by the ~340 000 identified amino acid residues (Figure 3.13B; yellow). On the other hand, hydrolysis with Lys-C and Glu-C in tandem decreased the coverage by at least ~12 % irrespective of the buffer environment (Figure 3.13B; blue). Analysis of individual amino acid sequences from each proteolytic digest revealed that some peptides generated by alternative combinations of proteases originated from distinct segments of corresponding proteins. Several research groups have reported similar observations and suggested utilising alternative proteases in parallel to improve the sequence coverage of the proteome (Swaney et al., 2010; Tsiatsiani and Heck, 2015). Therefore, the previously acquired raw MS files were re-analysed to determine the cumulative sequence coverage of the proteomes, when individual samples are combined into three experimental groups (Figure 3.13B; dark grey):

1) Samples digested with Lys-C, or Lys-C and Glu-C in NH_4HCO_3
2) Peptides generated by cleavage with Lys-C, or Lys-C and Glu-C in IAP buffer
3) Samples produced by digestion with Lys-C or trypsin in NH_4HCO_3

As anticipated, irrespective of the combination of enzymes selected for the experiment, joint analysis of two data sets resulted in an approximately 1.5-fold increase in additionally identified amino acid residues (Figure 3.13B; dark grey). Interestingly, the proteome sequence coverage acquired by the aggregation of MS files of peptides digested with Lys-C alone, or with Lys-C and Glu-C in IAP was increased compared to the samples cleaved in ammonium bicarbonate and thus supported the decision to carry out proteolytic digestions in IAP buffer.

The improved sequence coverage achieved by cleaving the samples with multiple endoproteinases in parallel demonstrates the importance of introducing Glu-C into the workflow. As the yield of 6His-SUMO2^{T90K} conjugates is limited, it was not desirable to split the sample into two. Thus, peptides retained on the 30 kDa ultrafiltration membrane after Lys-C cleavage alone were subjected to an additional digestion with Glu-C to facilitate their efficient elution and thus improve the sequence coverage without the requirement to subdivide the material.

3.2.10. Removal of K-ε-GG-specific antibody fragments from the mixture of purified peptides

As explained in Chapter 3.1.1, the enrichment of SUMO remnant-containing branched peptides is fundamental to enable the identification of modified proteins and facilitate the global mapping of sumoylated lysines. A selective isolation of target peptides can be achieved by immunoaffinity chromatography with a monoclonal antibody specific for diglycine-modified lysines. To date, two K-ε-GG-specific antibodies are commercially available, with each being stored in a different condition: lyophilised (Lucerna) or conjugated to protein A agarose matrix (Cell Signaling Technology) (Xu et

al., 2010; Kim et al., 2011). The latter product was considered more suitable for incorporation into the SUMO^{KGG} workflow, as the agarose matrix can facilitate straightforward removal of the antibody from the solution of low abundance peptides. However, upon low-pH elution of diGly-Lys-containing peptides, a large proportion of the K-ε-GG-specific antibody fragments were also released from the resin. Surprisingly, these were not efficiently removed from the sample by conventional C18-based reversed-phase chromatography, which is widely implemented for the desalting of peptides prior to the LC-MSMS analysis, and would normally be expected to retain large polypeptides. Consequently, the excessive antibody contaminants in the sample compromised the detection of low abundance SUMO remnant-containing peptides. Similar findings were reported by Udeshi et al. (Udeshi et al., 2013b).

To reduce the impact of antibody-derived contamination on the MS-based identification of sumoylated lysines, the K-ε-GG-specific antibody was chemically cross-linked to the protein A agarose matrix using a freshly prepared solution of bis(sulfosuccinimidyl)suberate (BS3). The importance of this cross-linking procedure was examined by comparing the release of antibody when bound to the resin non-covalently (conjugated) or when cross-linked to the agarose beads (Figure 3.14). Both, untreated and cross-linked resins were subjected to the mock immunoaffinity chromatography procedure without the addition of target peptides. TFA (0.15 %, vol/vol) was added to mimic the process of peptide elution and residual unbound antibody was extracted by heating the resin in 1 × Laemmli buffer. The amount of antibody fragments in each condition was analysed by SDS-PAGE followed by Coomassie brilliant blue staining (Figure 3.14). In the absence of treatment to stabilise

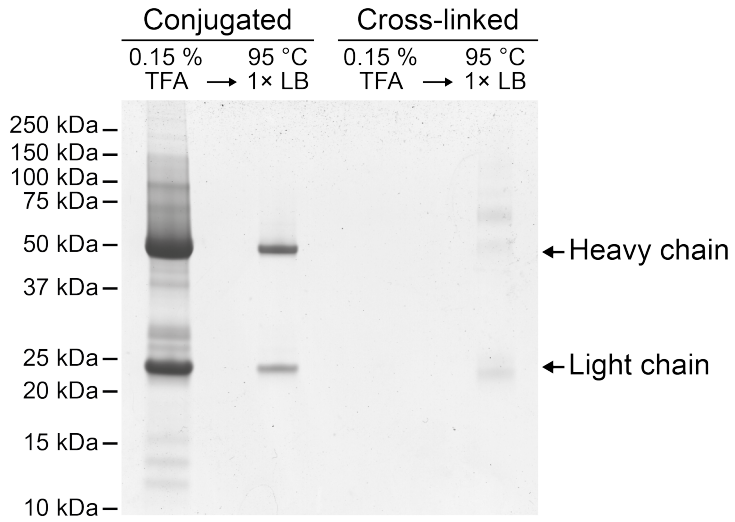

Figure 3.14 | Chemical cross-linking of K-ε-GG-specific antibody to protein A agarose beads reduces the release of antibody fragments upon low-pH elution.

The release of anti-K-ε-GG antibody from protein A agarose beads when either non-covalently bound (Conjugated; left) or cross-linked (right) was evaluated after incubating the resin in 0.15 % TFA. Residual unbound antibody was then extracted by heating the resin at 95 °C in 1 × Laemmli sample buffer (LB). The amount of antibody released with each procedure (0.15 % TFA; 95 °C 1 × LB) in each sample (Conjugated; cross-linked) was determined with Coomassie brilliant blue-stained SDS-PAGE protein gel.

the binding, a substantial proportion of antibody light and heavy chain fragments were extracted from the commercially supplied protein A resin (Figure 3.14; conjugated). In contrast, the amount of BS3-cross-linked K-ε-GG-specific antibody eluting with 0.15 % TFA was below the sensitivity threshold of the Coomassie brilliant blue stain and merely a marginal proportion was released upon heating the resin in strong detergent-based 1 × Laemmli buffer (Figure 3.14; cross-linked). Thus, BS3 can be utilised to stabilise K-ε-GG-specific antibody binding and prevent the excessive contamination of SUMO remnant-modified peptide samples. These findings are consistent with a study, where dimethyl pimelimidate (DMP) was used to cross-link the K-ε-GG-specific antibody to protein A beads (Udeshi et al., 2013b). However, in

current investigation, BS3 gave superior retention of the antibody on beads when compared to DMP.

3.2.11. Improving LC-MSMS settings for the optimal acquisition of low abundance peptides

After optimisation of the SUMO^{KGG} workflow (Figure 3.8), the settings of the LC-MSMS system were investigated. Two small-scale purification experiments were performed in parallel according to the established workflow. Specifically, the generated HEK293$^{6His\text{-}SUMO2\text{-}T90K}$ cells were heat-stressed at 43 °C for 30 minutes to enhance sumoylation and two portions of approximately 45 mg of cell lysate protein were subjected to nickel affinity chromatography. A small proportion (1 %) of each isolated fraction of 6His-SUMO2^{T90K} conjugates in solution was then cleaved directly with Lys-C, whereas majority of the substrates (99 %) were subsequently processed according to the SUMO^{KGG} workflow (Figure 3.8). Half of each sample was then analysed by LC-MSMS and peptide sequences were identified with the MaxQuant software package (Cox and Mann, 2008).

On average, ~4400 peptides corresponding to 1074 proteins were detected per experiment directly after nickel affinity chromatography (Figure 3.15). However, less than 0.07 % of them contained a SUMO remnant-modified lysine. Notably, the proportion of identified sumoylated peptides increased by more than 400-fold to nearly 30 % after the subsequent K-ε-GG-specific enrichment (Figure 3.15).

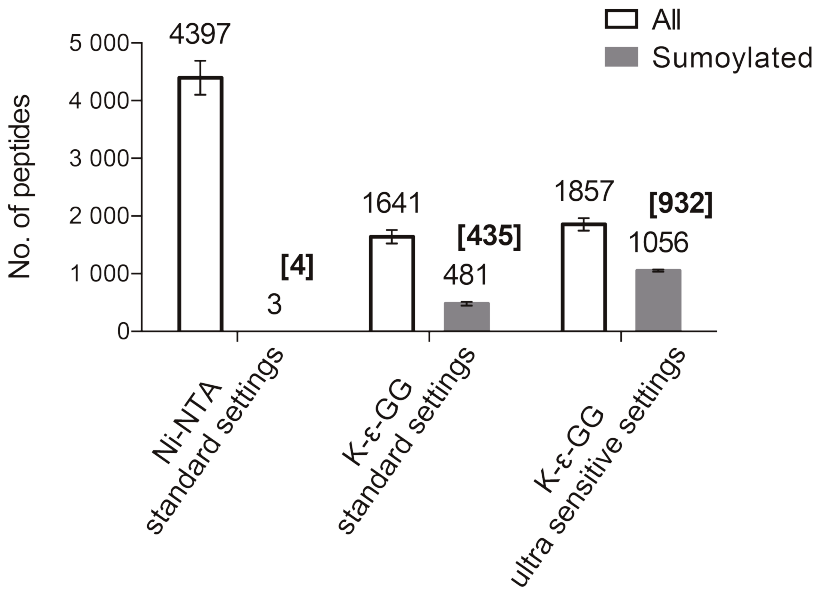

Figure 3.15 | Proteome-wide identification of sumoylation sites is enabled by the efficient enrichment and improved acquisition of modified peptides by MS.

Number of Gly-Gly-modified (grey) or all nonredudant peptide sequences (white) identified after nickel affinity chromatography using standard MS settings (left), or after subsequent diglycine-lysine (K-ε-GG)-specific immunoaffinity enrichment using standard MS parameters (middle), or MS settings refined for the optimal acquisition of low abundance peptides (right). The number of SUMO modification sites is reported in brackets. Data points are a mean of two replicate purifications from the same biological sample with standard deviations depicted as error bars.

Surprisingly, further assessment of the LC-MSMS performance during the analysis of diGly-Lys-enriched samples revealed that merely 8.2 % of the acquired MSMS spectra were matched to a specific amino acid sequence (Table 3.1). The concept of low sumoylation stoichiometry was reflected in multiple stages of the workflow. First, the yield of diGly-Lys-enriched peptides after the second purification stage was generally less than 1 μg, although ~45 mg of heat-stressed cell lysate protein was used as an input for the experiments. Furthermore, the inspection of LC-MSMS chromatograms

Table 3.1 | Summary statistics of raw files acquired with standard MS settings

No. of MSMS scans	23910[a]
Isotope patterns sequenced (%)	33.23[a]
PSMs (%)[b]	8.2[a]
Proportion of PSMs with IT <60 ms (%)[c]	0.42

[a] Average value; [b] Proportion of MSMS scans

[c] IT, ion injection time of the MSMS scan

revealed that the MS signal intensities of most peptide ions were relatively low, despite the extensive enrichment process. It was thus speculated that the majority of MSMS spectra remained unidentified due to the small amount of acquired precursor ions, which then produced poor fragment ion information. In support of this hypothesis, the median Andromeda score of all PSMs was only 53.6 (Figure 3.16) (Cox et al., 2011). Furthermore, nearly all ions (> 99.5 %), whose MSMS spectrum was matched to a sequence, were accumulated in the C-trap for the maximal allowed time of 60 ms (Table 3.1).

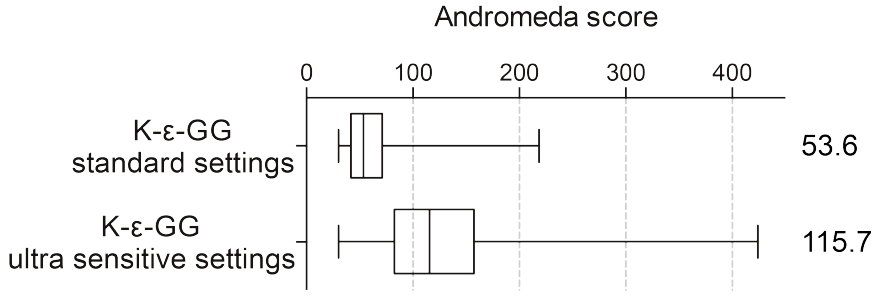

Figure 3.16 | Acquisition of LC-MSMS data with ultra sensitive instrument settings improves the quality of MSMS spectra.

Box plot illustrating the distributions of Andromeda scores of PSMs of MSMS spectra acquired with standard (top) or ultra sensitive (bottom) MS settings.

To address this issue, two solutions could be envisaged: either preparing and loading more sample onto the instrument, or investigating the LC-MSMS settings for the optimal acquisition of low abundance ions. It is often impractical to scale up experiments of this kind, especially considering the stoichiometry of protein sumoylation. Instead, the latter concept was examined and an ultra sensitive MS workflow was designed by altering mainly two parameters of the data-dependent MSMS acquisition:

1) The maximum injection time was increased from 60 ms to 1000 ms, to accumulate more precursor ions prior to their fragmentation and detection in Orbitrap.
2) Instead of ten, up to one precursor ion per full MS scan was selected for the data-dependent MSMS analysis to reduce the overall MS cycle time.

To assess the performance of this ultra sensitive MS workflow, identical amounts of previously described SUMO remnant-containing peptide samples were re-analysed by LC-MSMS. Notably, the experiment was performed on the same day with as few other variables as possible between individual runs to minimise systematic errors that could potentially arise from altered LC-MSMS instrumentation set-up or mass calibration. The impact of a longer ion injection time was clear, as more than twice as many diglycine-modified peptides were detected and efficiently sequenced from the same sample (Figure 3.15). Particularly, acquisition of MSMS spectra with settings optimised for the detection of low abundance ions enabled the identification of 1857 peptides per experiment and nearly 57 % of these sequences contained a site of Gly-Gly modification (Figure 3.15).

Furthermore, utilisation of ultra sensitive MS settings also facilitated the acquisition of better quality MSMS data and thus increased the reliability of the identified sequences. For instance, a fragmentation spectrum of a diGly-modified peptide of a human DNA topoisomerase 2-α was acquired with both, standard and optimised MS settings (Figure 3.17). However, advanced inspection of the two spectra revealed that the fragment ion coverage of the peptide was significantly improved upon the utilisation of ultra sensitive MS parameters (Figure 3.17; bottom panel). The higher quality of the MSMS spectra was also apparent in the Andromeda scores, as the median score of PSMs identified and detected with optimised MS settings increased from 53.6 to 115.7 (Figure 3.16).

The SUMO^{KGG} workflow is thus suitable for the high-throughput mapping of sites of sumoylation. In particular, two small-scale experiments with ~45 mg of heat-stressed HEK293$^{6His\text{-}SUMO2\text{-}T90K}$ cell lysate protein enabled the identification of 1468 diGly-modified lysines from 784 proteins (Figure 3.18). Although the majority (83.1 %) of these sumoylated lysines were mapped by utilising MS acquisition settings optimised for the detetion of low abundance ions, the remaining 16.9 % were identified by standard LC-MSMS analysis alone. Surprisingly, only 9.3 % of the detected isotope patterns were sequenced by tandem MS during the ultra sensitive LC-MSMS analysis (Table 3.2), which demonstrates the incomplete efficiency of the method to provide deep sequencing and could thus explain the absence of some sumoylated peptides. On the other hand, as the data acquired with standard MS settings was generally less reliable, a proportion of these uniquely identified modification sites could also arise from false peptide-to-spectrum matching.

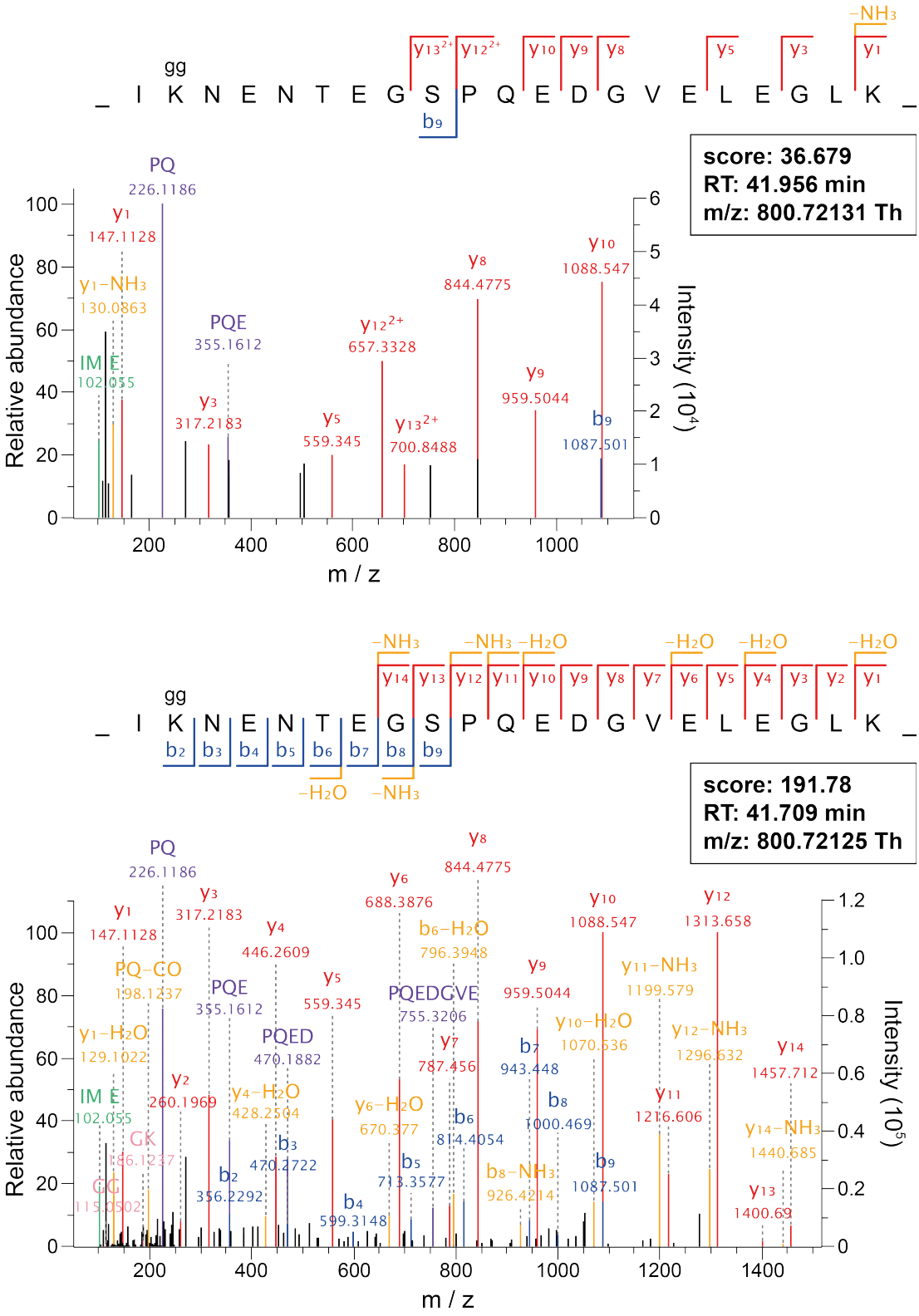

Figure 3.17 | Acquisition of MSMS spectra with ultra sensitive instrument settings increases the number of precursor ions injected into the Orbitrap, and improves peptide sequence coverage.

››

‹‹ **Figure 3.17 | Acquisition of MSMS spectra with ultra sensitive instrument settings increases the number of precursor ions injected into the Orbitrap, and improves peptide sequence coverage.**

Representative annotated MSMS spectra of a Gly-Gly remnant-containing peptide of DNA topoisomerase 2-α acquired by MS with standard (top) or ultrasensitive instrument settings (bottom). Peptide acquisition with refined ultrasensitive MS settings results in an increase in Andromeda score from 36.679 to 191.78 with a posterior error probability of 5.05×10^{-29}. Fragment ions extending from the N- or C-terminus of the peptide are named as b- (dark blue) or y-ions (red), respectively. Peptide internal fragments, ions with a loss of neutral molecule or immonium ions are illustrated in purple, yellow or green, respectively. Ions diagnostic for diGly modification are shown in pink. RT stands for retention time.

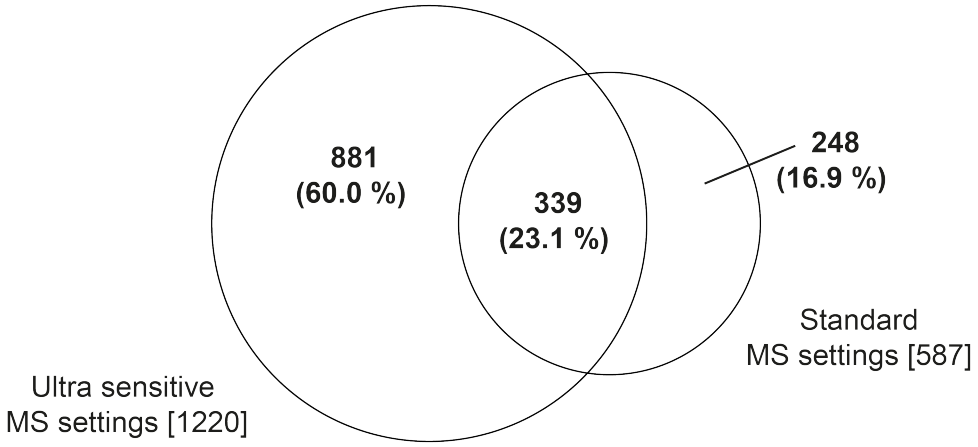

Figure 3.18 | The majority of the 6His-SUMO2^{T90K} modification sites are identified using ultra sensitive MS settings optimised for the acquisition of low abundance peptides.

A Venn diagram demonstrating the overlap in the identified sumoylated lysines between samples acquired using either standard or optimised ultra sensitive MS settings. The total number of sites mapped by either of the LC-MSMS analyses is displayed in brackets.

Table 3.2 | Summary statistics of raw files acquired with ultra sensitive MS settings

No. of MSMS scans	3716[a]
Isotope patterns sequenced (%)	9.34[a]
PSMs (%)[b]	38.52[a]

[a] Average value; [b] Proportion of MSMS scans

These observations demonstrate the requirement to optimise LC-MSMS settings based on the properties of individual samples. For instance, the critical characteristics of SUMO-containing samples, such as the abundance, dynamic range or diversity of substrates, is often inconsistent between samples that examine distinct biological phenomenons. Consequently, when the objective is to analyse SUMO remnant-modified branched peptides, it is recommended to follow these guidelines of LC-MSMS analysis:

a) A preliminary survey analysis should be performed with a small proportion (~10 %) of diGly-Lys-enriched peptide sample using standard LC-MSMS settings.
b) The complexity of the sample should be characterised based on the total ion current (TIC) and base peak LC chromatograms. If required, the length and the configuration of the linear gradient should be optimised to achieve an evenly dispersed elution of peptides from the reversed-phase column with an acceptable chromatographic peak shape.
c) The performance of the initial LC-MSMS run should be assessed based on the summary statistics of the processed raw MS files.
d) The settings of the subsequent data-dependent MSMS analyses should be altered to achieve a balance between ion injection time and the loop count.

Longer ion injection time improves the likelihood of peptide identification, but also increases overall MS cycle time. If the cycle time exceeds peptide elution time, the precursor ion selected for MSMS might no longer be present. Decreasing the loop count will decrease the MS cycle time, but should also be adjusted according to the number of co-eluting peptides to ensure that the precursor ions of most peptides are selected for fragmentation.

e) Given that long ion injection times are often required, one could consider increasing the transient length and improve the resolving power of the mass detection, provided that by doing so the overall MS cycle time is not increased excessively.

These guidelines were regularly adopted when employing the SUMO^{KGG} workflow to obtain high-quality LC-MSMS data for the global analysis of SUMO2 modification sites.

3.2.12. Uncovering the false discovery rate of the SUMO^{KGG} workflow

SUMO^{KGG} workflow relies on the assumption that endoproteinase Lys-C has an absolute specificity towards lysines (Raijmakers et al., 2010). Such a high substrate specificity is of utmost importance, as the proteolysis of peptide bonds N-terminal to the Gly-Gly motif at the C-termini of other human Ubls (Figure 3.1) can promote the formation of diglycine remnants with diverse origin, which can then be incorrectly assigned as the sites of 6His-SUMO2^{T90K}. As illustrated in Chapter 3.2.8, Lys-C is remarkably specific in IAP buffer (Figure 3.12A). However, despite the apparent low proportion of arginines (< 1.5 %) at the C-termini of identified peptide sequences

(Figure 3.12A; dark blue), their occurence is higher than expected based on the natural distribution of amino acids in a human proteome. Three members of the Ubl family (Ubiquitin, ISG15 and NEDD8) contain an arginine residue at the equivalent position to the T90K in 6His-SUMO2^{T90K} (Figure 3.1). Therefore, the proportion of identified Gly-Gly-lysines that are the bona fide targets of 6His-SUMO2^{T90K} modification should be determined.

To establish the true false discovery rate (FDR) of the SUMO^{KGG} workflow, the substrate specificity of Lys-C was evaluated by identifying the incorrectly assigned sites of Gly-Gly modification produced by unspecific Lys-C digestion (Figure 3.19). Thus, two cultures of HEK293 cells stably expressing either 6His-SUMO2^{T90K} or 6His-SUMO2 were incubated in a growth medium supplemented with either natural L-lysine ($^{12}C_6$, $^{14}N_2$; K0) and L-arginine ($^{12}C_6$, $^{14}N_4$; R0), or with the heavy stable isotope-containing L-lysine ($^{13}C_6$, $^{15}N_2$; K8) and L-arginine ($^{13}C_6$, $^{15}N_4$; R10), respectively. Both metabolically-labelled cultures were then incubated for 30 minutes at 43 °C to promote stress-induced protein sumoylation. After harvesting, equivalent quantities of both cultures were mixed based on the weight of the cell pellets. The combined sample was then processed according to the SUMO^{KGG} workflow and Gly-Gly-modified lysines were identified by LC-MSMS (Figure 3.19).

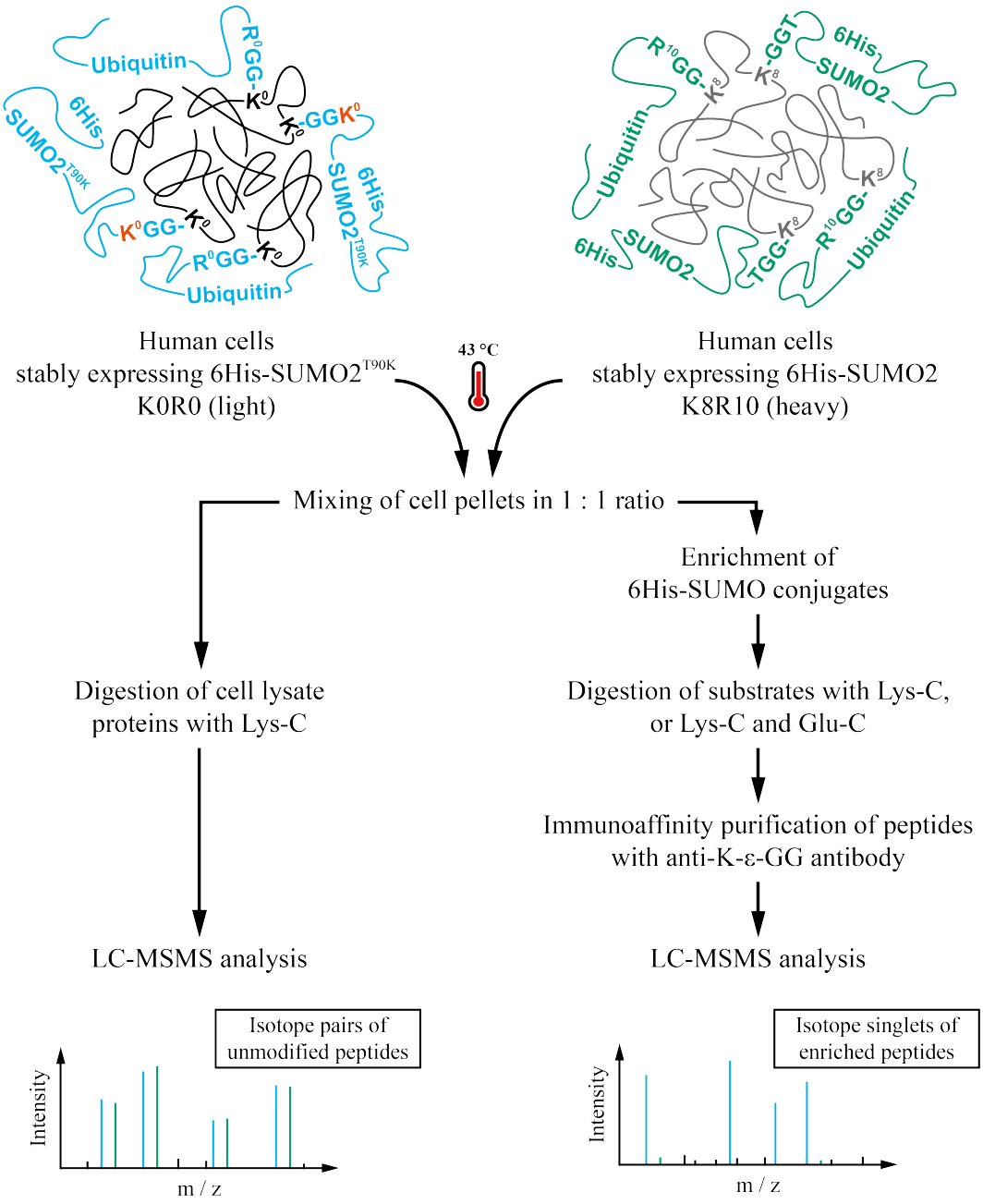

Figure 3.19 | Experimental strategy to understand the FDR of the SUMO^{KGG} workflow.

HEK293$^{6His\text{-}SUMO2\text{-}T90K}$ and HEK293$^{6His\text{-}SUMO2}$ cells were cultured in growth media supplemented with natural (K0R0) or heavy stable isotope-labelled (K8R10) L-lysine and L-arginine, respectively, and equivalent amounts of heat-treated cells were mixed. After lysis, ~5 % of protein was subjected to in solution digestion with Lys-C, while most of the protein mixture was processed using SUMO^{KGG} workflow. Peptides corresponding to total cell lysate or after diGly-Lys-specific enrichment were analysed by LC-MSMS. The false discovery rate (FDR) of the method was determined by measuring the relative abundances of diGly remnant-containing peptides in human cells expressing either 6His-SUMO2 or 6His-SUMO2^{T90K}.

If Lys-C catalyses the hydrolysis of peptide bonds C-terminal to lysine residues alone (Raijmakers et al., 2010), one should not detect any Gly-Gly remnant-containing peptides from K8R10-labelled sample expressing 6His-tagged wild-type SUMO2. However, the occurence of SILAC ratios can define the Gly-Gly modification sites that arise from incorrect cleavage and thus originate from other human Ubls (Figure 3.19). In total, 689 diglycine-modified lysines were identified (Figure 3.20). Remarkably, only two modification sites were detected from K8R10-labelled 6His-SUMO2-expressing cells with less than 21-fold difference compared to HEK293$^{6His\text{-}SUMO2\text{-}T90K}$ cells (Figure 3.20; Table 3.3). Interestingly, both of these falsely assigned sites occured on Ubiquitin. However, the 6His-SUMO2-to-6His-SUMO2^{T90K} ratio of K11-containing peptide of Ubiquitin was noticeably small and thus suggested that this lysine can act as a substrate to multiple Ubls (Table 3.3).

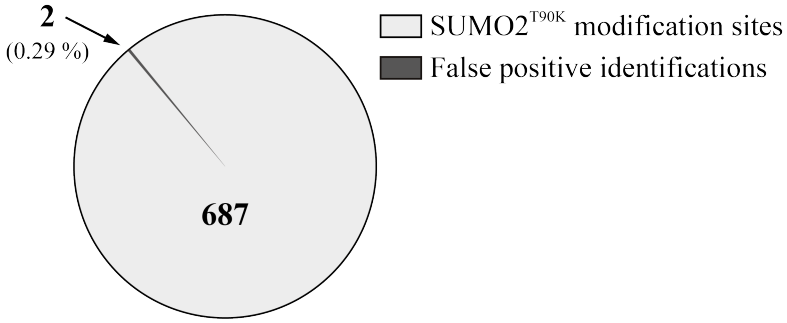

Figure 3.20 | Less than 1 % of the Gly-Gly-modified peptides identified using the SUMO^{KGG} workflow arise from proteins conjugated to a different Ubl.

A pie chart demonstrating the proportion of diGly remnant-containing peptides identified from 6His-SUMO2^{T90K}-expressing cells only (SUMO2^{T90K} modification sites; light grey) or also from those expressing 6His-SUMO2 (False positive identifications; dark grey). The false discovery rate (FDR) of the SUMO^{KGG} workflow is ~0.29 %.

Table 3.3 | False positive identifications

Seq. no.	Protein name	Position of modified lysine	6His-SUMO2 / 6His-SUMO2^{T90K} (normalised)	Fold difference
1	Ubiquitin	K48	0.93814	1.1
2		K11	0.04758	21.0

In conclusion, as indicated by the very low FDR of the SUMO^{KGG} workflow (< 1.0 %), this highly reliable biochemical approach can be utilised for the global identification of Gly-Gly modification sites that originate solely from 6His-SUMO2^{T90K} (Figure 3.20).

3.2.13. Benchmarking the SUMO^{KGG} workflow

First trial experiments with the HEK293 cells stably expressing 6His-SUMO2^{T90K} according to the SUMO^{KGG} workflow enabled the identification of a few hundred SUMO2 modification sites. The extensive optimisation process described throughout this chapter improved the method significantly and ~1000 sumoylated lysines are now routely identified from ~1.75×10^8 HEK293$^{6His\text{-}SUMO2\text{-}T90K}$ cells corresponding to ~22.5 mg of cell lysate protein (Figure 3.15).

To estimate the reproducibility of the SUMO^{KGG} workflow, duplicate cultures of HEK293$^{6His\text{-}SUMO2\text{-}T90K}$ cells were processed in parallel according to the SUMO^{KGG} workflow and analysed using the ultra sensitive LC-MSMS settings. In total, 1220 sumoylation sites were identified and 52.7 % of them were common between the replicate purifications (Figure 3.21). However, despite adequate reproducibility, one should consider performing multiple replicate experiments or LC-MSMS runs to acquire a deep coverage of the modification sites of the sumoylated proteome.

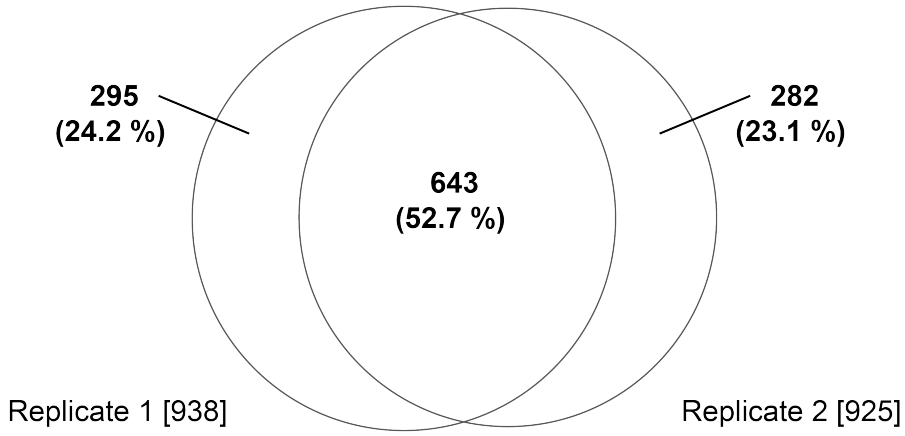

Figure 3.21 | Reproducibility of the SUMO^{KGG} workflow.

A Venn diagram illustrating the overlap in the identified SUMO modification sites between two replicate purifications of 6His-SUMO2^{T90K}-modified proteins and corresponding Gly-Gly remnant-containing peptides. The values in brackets show the total number of sites identified in each sample.

A combined analysis of 18 LC-MSMS raw files acquired during the development stage of the SUMO^{KGG} workflow produced a comprehensive list of 2034 SUMO2 modification sites from 990 proteins extracted from heat-stressed HEK293$^{6His\text{-}SUMO2\text{-}T90K}$ cells (Supplementary table S1; Appendix 6.1). Approximately 43.3 % of the modified lysines were identified only from samples processed additionally by Glu-C digestion. Furthermore, the majority of MSMS fragmentation spectra of diGly-modified branched peptides were manually validated according to the following stringent criteria:

a) Broad coverage of b and y fragment ion series extending from the N- or C-terminus of the peptide, respectively.
b) Extensive identification rate of intensive fragment ion peaks.
c) Precursor ion mass error less than 2 parts per million (ppm) after mass recalibration or 4 ppm in case of unsuccessful recalibration.

d) Preferential fragmentation N-terminal to proline or C-terminal to glutamic and aspartic acids.

e) Existence of ions diagnostic for diGly modification (GG^+, m/z +115.0505 Th; KGG^+, m/z +186.1237 Th).

f) Presence of an ion series with a neutral loss corresponding to a single glycine (m/z 57.0215 Th).

Moreover, the localisation probability of all modified lysines was confirmed to be greater than 75 % and peptides identified based on the reverse decoy database were removed. Finally, identification of multiple peptide sequences containing the same Gly-Gly-modified lysine was considered beneficial for the validity of the assigned SUMO2 modification site.

Overall, this list of 2034 SUMO2 modification sites was ~20-fold larger than the biggest site-specific study published prior to this research (Figure 3.22) (Matic et al., 2010). In addition, the current study increased the number of all known sites of sumoylation determined by MS-based methods by more than 12-fold (Hornbeck et al., 2012).

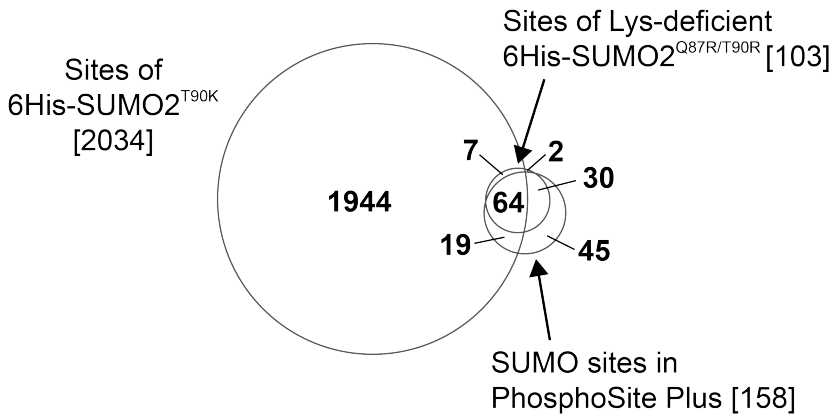

Figure 3.22 | The SUMO^{KGG} workflow increases the known number of sumoylation sites by more than tenfold.

A Venn diagram showing the comparison among SUMO2 modification sites identified in this study and two data sets published prior to the completion of this study: (1) an MS-based study using Lys-deficient 6His-SUMO2$^{Q87R/T90R}$-expressing HeLa cells (Matic et al., 2010), (2) all other studies using MS as annotated in the PhosphoSitePlus database as of January 2014 (Hornbeck et al., 2012). The total numbers of sumoylation sites are shown in brackets.

In summary, these improvements in the detection and identification of sumoylated peptides were achieved through number of technical and methodological enhancements, including the use of T90K version of SUMO2 and K-ε-GG-specific antibody, chemical cross-linking of the antibody, utilisation of additional in-tandem digestion with Glu-C, and optimisation of the MSMS acquisition conditions.

3.3. Discussion

3.3.1. SUMO^{KGG} workflow for the global identification of SUMO2 modification sites

Prior to the development of the SUMO^{KGG} workflow, experimental strategies utilised for the identification of sites of sumoylation in human cells employed overexpression of SUMO with an amino-terminal affinity tag and mutations in the C-terminal region to shorten the sequence of the remnant peptide after trypsin digestion (Blomster et al., 2010; Matic et al., 2010; Galisson et al., 2011; Lamoliatte et al., 2013; Schimmel et al., 2014). Isolation of sumoylated proteins was thus achieved through an affinity purification. However, lack of enrichment procedures specific for modified peptides hindered the MS-based detection and identification of low abundance SUMO remnant-containing sequences. Two previous studies acknowledged the requirement for efficient purification of modified peptides, but the applied approaches enabled the mapping of a maximum of 103 sumoylated lysines only (Blomster et al., 2010; Matic et al., 2010). This number, however, was at least tenfold less than expected based on the amount of proteins published as SUMO substrates (Golebiowski et al., 2009; Bruderer et al., 2011; Tatham et al., 2011; Becker et al., 2013). Alternatively, sophisticated bioinformatic tools were developed to remove the requirement of C-terminal mutations by facilitating database search engine-based identification of chimeric MSMS spectra of large branched peptides (Pedrioli et al., 2006; Hsiao et al., 2009; Lamoliatte et al., 2013). However, these tools had little success with complex peptide mixtures of human proteome.

In this chapter, a SUMO^{KGG} workflow was developed to enable global, high-throughput, and unbiased identification of SUMO2 modification sites from complex samples (Figure 3.8). A stable HEK293 cell line expressing 6His-SUMO2^{T90K} was generated to act as an *in vivo* model for sumoylation. The enrichment of sumoylated proteins followed by the affinity purification of SUMO Gly-Gly remnant-containing peptides after Lys-C digestion provided material from which modification sites could be identified. Notably, nearly all Gly-Gly-lysines produced through Lys-C proteolysis originated from 6His-SUMO2^{T90K}, as demonstrated by ~0.29 % false discovery rate of the workflow (Figure 3.20). A number of steps of the SUMO^{KGG} workflow were critical to this enhanced sensitivity, efficiency and robustness. As anticipated, a major improvement in sensitivity was achieved through the use of the antibody specific for diGly-lysines. This increased the proportion of SUMO remnant-containing peptides in the final samples by more than 430-fold (Figure 3.15). The sensitivity of the SUMO^{KGG} workflow was also improved by altering MS settings for the optimal acquisition of low abundance diGly-Lys-containing peptide ions. Improved coverage of the human proteome was achieved by the introduction of an additional cleavage step using endoproteinase Glu-C, which nearly doubled the number of identified sumoylation sites. Streamlining sample handling increased the yield and robustness of the SUMO^{KGG} workflow. All abovementioned enhancements thus supported the development of a workflow that can be utilised to routinely identify ~1000 sumoylated lysines from ~1.75×10^8 heat-stressed HEK293$^{6His\text{-}SUMO\text{-}T90K}$ cells. In conclusion, the SUMO^{KGG} workflow was the first method to enable high-throughput identification of the sites of sumoylation and demonstrated the depth of the human proteome sequence that can be conjugated to SUMO2 (Appendix 6.2) (Tammsalu et al., 2014; Tammsalu et al., 2015).

3.3.2. Additional applications of the SUMO^{KGG} workflow

A common feature of Ubls is the C-terminal Gly-Gly motif (Flotho and Melchior, 2013). The developed SUMO^{KGG} workflow could, in principle, be adapted for the global identification of the sites of any Ubl. Many key questions of the biology of the Ubl family proteins could thus be addressed with this workflow. Such as, which lysines are unique or which are shared among Ubls? Is there a site specificity in the Ubl conjugation in response to cellular stresses? Moreover, current SUMO^{KGG} workflow could be tailored to compare the relative abundances of distinct Ubls attached to the same target lysine and measure the extent of site-specific Ubiquitin-like protein crosstalk.

The C-terminal diGly motif of the mature SUMO molecule is also conserved across species (Saitoh and Hinchey, 2000). The developed SUMO^{KGG} workflow could thus be applied to map the sites of SUMO modification in other model organisms. In collaboration with a research group led by Genevieve Thon (University of Copenhagen), the workflow was utilised for the global identification of SUMO modification sites in *S. pombe* (Appendix 6.2) (Kohler et al., 2015). In this study, the single gene encoding SUMO in fission yeast (*pmt3*) was replaced with 6His-SUMO^{L109K} to generate an organism compatible with the SUMO^{KGG} workflow. Thus, the 6His-SUMO^{L109K} was expressed from the endogenous locus and served as the sole source of cellular SUMO. Unlike the severe growth defects of SUMO-deficient *S. pombe* (Δ*pmt3*) (Tanaka et al., 1999), 6His-SUMO^{L109K} did not compromise the survival nor growth rate of the fission yeast (Figure 3.23A and B). With MS-based quantitative proteomics it was shown that no apparent proteome-level compensation was required

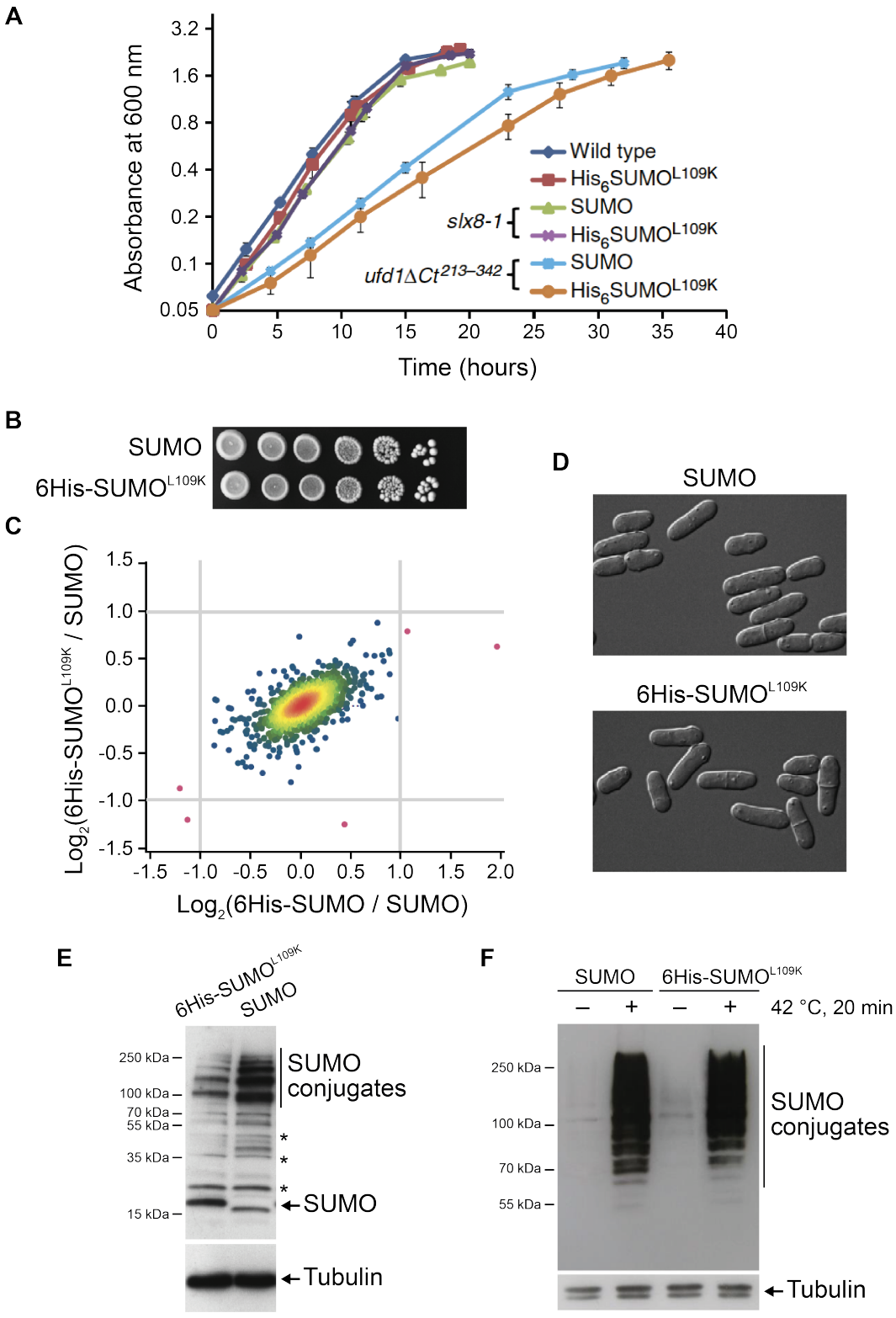

Figure 3.23 | The phenotype of *S. pombe* strain expressing solely 6His-SUMO^{L109K} is indistinguishable from wild-type counterpart.

››

‹‹ **Figure 3.23 | The phenotype of *S. pombe* strain expressing solely 6His-SUMO^{L109K} is indistinguishable from wild-type counterpart.**

(**A**–**B**) Growth curves (A; dark blue and red) or tenfold dilution series (B) of *S. pombe* strains expressing either SUMO or 6His-SUMO^{L109K}. (**C**) Scatter plot illustrating the log-transformed (base 2) relative abundances of 2266 proteins between *S. pombe* strains expressing either SUMO or 6His-SUMO (abscissa), or 6His-SUMO^{L109K} (ordinate). (**D**) Micrographs of strains expressing either SUMO or 6His-SUMO^{L109K}. (**E**) Western blot images of cell lysate proteins of *S. pombe* strains expressing either SUMO or 6His-SUMO^{L109K} detected using antibodies recognising either SUMO (top) or Tubulin (bottom). Nonspecific immunoreactions are indicated by asterisks. (**F**) Western blot images of cell lysate proteins from unperturbed (–) or heat-stressed (+; 42 °C, 20 min) *S. pombe* strains expressing either SUMO or 6His-SUMO^{L109K} detected using antibodies recognising either SUMO (top) or Tubulin (bottom). Figure is adapted from (Kohler et al., 2015).

to achieve the observed phenotype (Figure 3.23C). Furthermore, the morphology of *S. pombe* cells expressing either SUMO or 6His-SUMO^{L109K} was indistinguishable (Figure 3.23D). Finally, Western blot analyses demonstrated that 6His-SUMO^{L109K} was efficiently conjugated to target proteins and responded to cellular stress in a similar manner to the wild-type counterpart (Figure 3.23E and F). Utilisation of the SUMO^{KGG} workflow in this context enabled the identification of 1028 SUMO modification sites from 468 fission yeast proteins. The system was also used to measure the effect on sumoylated proteome in *S. pombe* cells deficient in the function of SUMO-targeted E3 Ubiquitin-protein ligase subunit Slx8 or Ubiquitin fusion degradation protein 1 (Ufd1). This revealed that several telomere- and centromere-associated proteins of fission yeast were coordinately processed by the combined actions of the Cdc48-Ufd1-Npl4 segregase and SUMO-targeted E3 Ubiquitin-protein ligases Rfp1/Slx8 and Rfp2/Slx8 (STUbLs; Appendix 6.2) (Kohler et al., 2015). The SUMO^{KGG} workflow has also been applied for the global identification of sumoylated lysines from *C. elegans* germline and early embryos (Appendix 6.2) (Pelisch et al., 2016).

Although heat stressed HEK293$^{6His\text{-}SUMO2\text{-}T90K}$ cells were utilised during the development of SUMO^{KGG} workflow, the protocol could be used to investigate the sites of sumoylation involved in other cellular processes, such as cell division, DNA replication, transcription, or in cells exposed to various stress-inducing stimuli. For instance, SUMO^{KGG} workflow has been used to reveal the dynamic proportion of the more than 600 SUMO2 modification sites identified from HEK293$^{6His\text{-}SUMO2\text{-}T90K}$ cells after a treatment with a suspected Ubl protease inhibitor GRE010 (collaboration with the research group of Professor Angus Lamond).

3.3.3. Comparison of the SUMO^{KGG} workflow with alternative methods

In recent years, the ability to globally determine the sites of SUMO modification has been one of the central objectives of the field of protein sumoylation. Consequently, several independent research groups have focused on developing methods for the high throughput identification of modified lysines. Since the publication of the SUMO^{KGG} workflow, four additional protocols enabling the mapping of more than a hundred sumoylation sites have thus been developed (Hendriks et al., 2014; Impens et al., 2014; Lamoliatte et al., 2014; Hendriks et al., 2015a). The foundations of all these techniques, including the SUMO^{KGG} workflow, are surprisingly similar. For instance, all published protocols depend on an affinity-tagged variant of SUMO and most of them utilise both protein- and peptide-level enrichments. However, several key technical distinctions exist and should thus be taken into consideration, when analysing the archives of previously identified modification sites or selecting an appropriate approach for future experiments.

Similarly to the SUMO^{KGG} workflow, the method published by Impens et al. utilised a 6His-tagged variant of SUMO allowing the affinity purification of modified proteins (Impens et al., 2014). Furthermore, the residue preceding the C-terminal diGly sequence was mutated to enable the enrichment of proteolytically digested SUMO branched peptides with K-ε-GG-specific antibody. However, in contrast to the SUMO^{KGG} workflow, the residue was mutated to Arg and trypsin digestion was utilised to generate peptides. As more members of the Ubl-family contain an arginine in the equivalent position, additional measures had to be taken to discriminate between the sites of SUMO, Ubiquitin, NEDD8, and ISG15. This was accomplished by comparing the MS signal intensities of diGly-Lys-containing peptides derived from SILAC-labelled cells expressing either wild-type or engineered SUMO. Only peptides with an MS signal detected exclusively from mutant cells were acknowledged as genuine substrates (Impens et al., 2014). This type of analysis is prone to type II errors ("false negatives"), as lysines modified by various Ubls, including SUMO, could be incorrectly rejected. Furthermore, co-purification of modified peptides originating from multiple Ubls increases the complexity of samples, and can thus hamper the identification of low abundance sumoylated substrates. As a result, less than a third of the acquired peptides contained a SUMO-derived remnant sequence (Impens et al., 2014; Hendriks and Vertegaal, 2016a). On the other hand, this is the only method that to date has been applied to simultaneously analyse the sites of SUMO1 and SUMO3, identifying 295 or 167 modified lysines, respectively (Impens et al., 2014).

A 6His-SUMO construct was also utilised for the method developed by Lamoliatte et al. and the enrichment of sumoylated conjugates was achieved by nickel affinity chromatography (Lamoliatte et al., 2014). Moreover, an arginine mutation was

introduced to the C-terminal region of SUMO2 to create an additional cleavage site for trypsin. However, the mutation was introduced to the sixth position from the carboxyl-terminus of the mature molecule to distinguish between the tryptic remnants of SUMO and other Ubls. Additionally, Gln^{88} was replaced by an asparagine to potentially analyse the redundancy between the targets of SUMO2 and SUMO3, although the topic was not addressed. After trypsin digestion, the NQTGG-modified branched peptides from $SUMO2^{Q87R/Q88N}$ conjugates were enriched with a rabbit monoclonal antibody UMO 1-7-7, which to date is not available to the wider scientific community. Surprisingly, identification of the MSMS spectra of NQTGG-Lys-containing peptides required utilisation of a specialist software application developed in-house (Lamoliatte et al., 2014). In total, Lamoliatte et al. were able to identify 954 sites of SUMO2 modification in response to proteasome inhibition with MG132.

The third protocol was an optimised version of the most efficient method published prior to $SUMO^{KGG}$ workflow (Matic et al., 2010; Hendriks et al., 2014). As before, a mutant of SUMO2 completely lacking lysine residues was utilised to render the molecule resistant to Lys-C digestion and Gln^{87} was mutated to Arg to generate trypsin digested branched peptides with a shorter SUMO2 remnant (QQTGG). However, the 6His-tag was replaced with a decahistidine (10His) stretch to allow for more stringent affinity chromatography procedure, which generated samples of higher purity. Furthermore, the affinity purification was utilised twice: before Lys-C digestion to enrich for sumoylated proteins, and after proteolysis to isolate target peptides attached to Lys-C resistant $SUMO2^{K0/Q87R}$. After subsequent trypsin digestion, branched peptides were identified utilising new MS instrumentation. Altogether, these adjustments enabled the identification of more than 4000 10His-$SUMO2^{K0/Q87R}$ modification sites

from human cells exposed to various cellular stresses (Hendriks et al., 2014; Hendriks and Vertegaal, 2016b). Importantly, Lys-deficient 10His-SUMO2$^{K0/Q87R}$ cannot form polymeric SUMO chains, which can have a distinct function compared to mono-sumoylation (Tatham et al., 2008), and the phenotype and viability of cells with 10His-SUMO2$^{K0/Q87R}$ as their sole copy is yet to be addressed (Hendriks et al., 2014).

The most recent method, named as protease-reliant identification of SUMO modification or "PRISM", enabled sumoylation site identification using wild-type SUMO sequence (Hendriks et al., 2015a). Expression of exogenously introduced 10His-tagged SUMO was still essential to facilitate the affinity purification of conjugates. After enrichment, unmodified lysine residues of denatured purified proteins were acetylated using sulfosuccinimidyl acetate (sulfo-NHS-acetate) and conjugates were desumoylated with SENP2. Deconjugated substrate proteins were then digested with trypsin and analysed by LC-MSMS. Location of target sites was inferred indirectly via the identification of peptides containing unmodified lysines (Hendriks et al., 2015a). The major weakness of this approach is the reliance on the sulfo-NHS-acetate acetylation reaction to go to completion, as residual unmodified lysines could be misassigned as SUMO target sites. Furthermore, potential side reactions of sulfo-NHS-acetate can introduce unknown modifications and complicate the identification of peptide sequences. *In vitro*, SENP2 has been reported to deconjugate SUMO1, SUMO2 and SUMO3 (Nishida et al., 2001), which suggests that paralog-specific sites of sumoylation can only be determined utilising protein-level enrichment prior to SENP2 treatment. Therefore, lysines of (co-)purified substrates modified by a different SUMO paralog can be misassigned.

In practice, SUMO^{KGG} workflow is comparable to all abovementioned methods in terms of protocol duration (~5 versus 3–7 days) and the amount of cell culture required per LC-MSMS analysis (90 million versus 20–200 million cells) (Hendriks et al., 2014; Impens et al., 2014; Lamoliatte et al., 2014; Tammsalu et al., 2014; Hendriks et al., 2015a; Tammsalu et al., 2015; Hendriks and Vertegaal, 2016b). Moreover, the number of identified SUMO2 modification sites is proportionate with the highest performer among the other methods (2034 versus 2307 sites after heat stress) (Hendriks et al., 2014). However, as illustrated by the large proportion of Gly-Gly-Lys-containing peptides present in enriched samples (Figure 3.15), the purification efficiency of the SUMO^{KGG} workflow is superior to all other techniques (~60 % versus 2–30 %) (Tammsalu et al., 2014; Hendriks and Vertegaal, 2016a). Furthermore, the SUMO^{KGG} workflow utilises a form of SUMO with a single amino acid substitution near the C-terminus, thus reducing the risk of non-wild-type behaviour of the SUMO protein. Mutating the residue preceding the diGly motif to lysine and using Lys-C assures that all modified peptides originate exclusively from SUMO2^{T90K}. Moreover, the relatively small SUMO-derived Gly-Gly modification on target lysines is identifiable by standard database search engines, thus preventing the requirement for specialist software applications or add-ons.

3.3.4. Future directions of SUMO site proteomics

Hitherto, all available methods for the high-throughput identification of sumoylation sites rely on the expression of exogenously introduced versions of SUMO paralogs. Ultimately, mapping of lysines modified by endogenous SUMO would be the ideal, as the use of exogenous or mutated forms of the protein would be avoided and the

approach could be applied to tissues or clinical samples. The ultimate incarnation of SUMO proteomics is a method that allows the identification and quantification of the sites of endogenous sumoylation, while distinguishing between the three paralogs.

Despite the substantial advancements in LC-MSMS instrumentation, with present technology the requirement to enrich for low abundance sumoylated molecules is inevitable. However, a methodology for the purification of endogenous SUMO remnant-containing peptides after proteolysis is lacking. To date, the majority of sumoylation site analyses have focused on the identification of SUMO remnant-modified lysines after trypsin digestion, and required SUMO mutations to allow peptide-specific enrichment (Hsiao et al., 2009; Blomster et al., 2010; Matic et al., 2010; Lamoliatte et al., 2013). However, investigation into the suitability of alternative proteases has been limited. Thus, a comprehensive assessment of biochemical properties of SUMO remnant-containing branched peptides after proteolytic digestions with various endopeptidases either alone or in combination would be beneficial. This type of analysis could reveal SUMO branch-specific features, such as suitable fixed size or chemical characteristics that could enable their efficient separation from unmodified peptides. Alternatively, these branched peptides could serve as an epitope to generate novel SUMO remnant-specific antibodies.

Recently, the application of a wild-type α-lytic protease (WaLP) in bottom-up proteomics experiments was demonstrated (Silen et al., 1989; Meyer et al., 2014). WaLP cleaves peptide bonds predominantly from the C-terminal side of Val, Ala, Thr, and Ser. Therefore, the protease could be utilised to generate Gly-Gly-Lys-containing peptides from endogenous SUMO conjugates. However, this enzyme cannot be

utilised to map paralog-specific sites of modification, as all SUMO molecules contain Thr in the equivalent position (Figure 3.1). Furthermore, with lower affinity, WaLP digests peptides C-terminal to Gly, Leu and Met (Meyer et al., 2014). Ubiquitin-like protein FUBI contains leucine in the equivalent position to Thr^{90} in SUMO, and WaLP-generated Gly-Gly-Lys-containing peptides could thus originate from substrate proteins modified by multiple Ubls. Moreover, WaLP-specificity towards peptide bonds C-terminal to glycines could shorten the Gly-Gly remnant and thus render the branched peptides unsuitable for the subsequent enrichment with K-ε-GG-specific antibody. However, despite all these limitations, WaLP-based method could validate the sites of SUMO modification identified with currently available approaches. But until such a method becomes available, techniques like the $SUMO^{KGG}$ workflow can be utilised to study sumoylated lysines in different organisms, cell types, and stress conditions.

4.

Global reprogramming of SUMO signalling in response to heat-induced proteotoxic stress

4.1. Introduction

Protein posttranslational modification by SUMO2/3 is enhanced in cells subjected to various protein-damaging stimuli, such as oxidative stress, addition of ethanol, inhibition of protein degradation, or exposure to severe temperature fluctuations (Saitoh and Hinchey, 2000). The dynamics of sumoylation appears to correlate with the accumulation of mis- and unfolded proteins. For instance, conjugation to SUMO2/3 in response to proteasome inhibition depends on the accumulation of newly synthesised misfolded polypeptides (Tatham et al., 2011). Exposure to elevated temperature, however, triggers a fast translation-independent substrate sumoylation, which correlates with the rapid and widespread aggregation of denatured proteins (Tatham et al., 2011; Wallace et al., 2015). Importantly, conjugation to SUMO2/3 promotes the survival of cells after a severe heat shock (Golebiowski et al., 2009).

Hundreds of proteins are sumoylated in response to hyperthermic cellular stress and many of them can bind chromatin (Golebiowski et al., 2009; Niskanen et al., 2015; Seifert et al., 2015). In particular, an exposure to elevated temperature induces a rapid SUMO2 conjugation to proteins associated with promoters and enhancers of active genes. Modification of these chromatin-binding proteins appears to require the transcriptional activity of target genes (Niskanen et al., 2015; Seifert et al., 2015). Similarly to the heat shock response in general, sumoylation of chromatin-associated proteins is induced by the accumulation of either un- or misfolded proteins. (Seifert et al., 2015). However, the general function of SUMO2 recruitment to chromatin is still unclear, as contradictory information has been reported stating that sumoylation either

maintains or restricts the expression of target genes (Niskanen et al., 2015; Seifert et al., 2015).

Subcellular fractionation-based studies have revealed that the unconjugated form of SUMO2/3 is preserved in the cytoplasm, whereas sumoylated substrates are almost exclusively nuclear. After the exposure to protein-damaging stimuli, the majority of SUMO-bound conjugates are associated with the detergent-insoluble fraction upon cell lysis (Tatham et al., 2011; Seifert et al., 2015). Interestingly, this insoluble material appears to consist of large protein complexes that are tightly associated with DNA. Preliminary experiments have suggested that multiple components of these protein complexes could act as substrates of SUMO2 modification (Seifert et al., 2015).

The aim of this project was to uncover the global dynamics of SUMO2 signalling in heat-stressed cells by providing an extensive quantitative resource of sumoylated proteins together with the exact SUMO2 modification sites. A direct quantitative comparison of normal and heat-stressed cells would enable the verification of the major targets of proteotoxic stress-induced sumoylation by demonstrating the extent of modification of individual substrates in each condition. The generated comprehensive knowledgebase of SUMO2 modification sites could also be utilised by the scientific community to uncover the physiological functions of sumoylation.

Considering that the predominantly cytoplasmic pool of unconjugated SUMO2 decreases rapidly upon heat stress, and the emerging sumoylated conjugates are mainly nuclear, it was anticipated that many other proteins might be under similar proteotoxic stress-induced spatial regulation (Tatham et al., 2011; Seifert et al., 2015).

Surprisingly, the global changes to the spatial distribution of the human proteome in response to hyperthermic stress have not been addressed.

To this end, a quantitative SILAC (stable isotope labelling by amino acids in a cell culture)-based workflow was devised (Ong et al., 2002). The developed method was designed to enable direct comparison between individual protein abundances in the cytoplasmic or nuclear fraction of either normal or heat-stressed HEK293$^{6His\text{-}SUMO2\text{-}T90K}$ cells. Furthermore, the SUMO^{KGG} workflow was utilised to identify the sites of SUMO2 modification and to determine the extent of sumoylation of each substrate in unperturbed cells or after the exposure to heat stress.

4.2. Results

4.2.1. Time course of 6His-SUMO2^{T90K} conjugation to substrate proteins in response to hyperthermic stress

In response to heat-activated cellular stress, available SUMO2 is covalently attached to substrate proteins (Saitoh and Hinchey, 2000; Golebiowski et al., 2009). To determine the dynamics of heat-induced sumoylation, HEK293$^{6His\text{-}SUMO2\text{-}T90K}$ cells were exposed to 0, 15, 30, 60, or 120 minutes of hyperthermic stress at 43 °C. Cell pellets were disrupted in a buffer containing 100 mM 2-chloroacetamide to inhibit desumoylation (Bruderer et al., 2011), and proteins were segregated into nuclear and cytoplasmic fractions. Equivalent amount of each subcellular fraction (protein wt/wt) was then loaded onto SDS-PAGE gels and subjected to Western blot analyses for the detection of SUMO2/3, 6His, Tubulin, or Lamin A/C (Figure 4.1).

Immunoblot images of SUMO2/3 and 6His indicated that the conjugation of SUMO2/3 to target proteins was maximal after 30–60 minutes of heat stress (Figure 4.1A and B). Consistent with previously published observations, the sumoylated substrates resolved in the high-molecular-weight area of the protein gel and were predominantly detected in the nuclear fraction (Tatham et al., 2011; Seifert et al., 2015). Meanwhile, the majority of unconjugated SUMO2 monomers were found in the cytoplasmic fraction (Figure 4.1A and B).

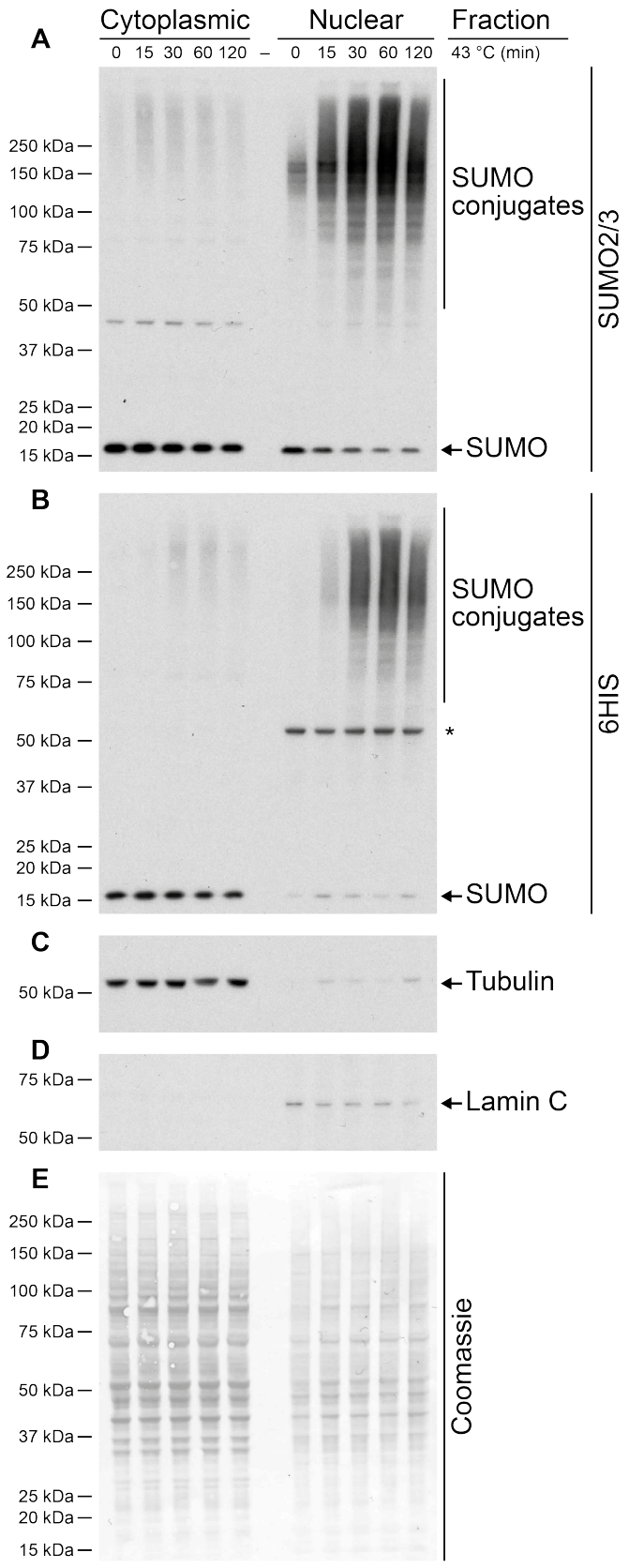

Figure 4.1 | The apex of 6His-SUMO2^{T90K} conjugation is reached in 30–60 minutes of exposure to acute heat stress. HEK293$^{6His\text{-}SUMO2\text{-}T90K}$ cells cultured at 37 °C were shifted to 43 °C and cells were harvested either before (0 min) or after 15, 30, 60, and 120 minutes of treatment. Proteins from each sample were separated into cytoplasmic and nuclear fractions, and equivalent amounts were subjected to Western blot analyses for the detection of: (**A**) SUMO2/3, (**B**) 6His, (**C**) Tubulin, and (**D**) Lamin A/C. (**E**) Coomassie brilliant blue-stained PVDF membrane validating the equivalent loading and consistent transfer of proteins.

Western blot images of Tubulin and Lamin A/C demonstrated the efficiency of the protein separation and served as loading controls for the corresponding fractions (Figure 4.1C and D). As expected, Tubulin was detected predominantly in the cytoplasmic fraction (Figure 4.1C), and Lamin A/C was found exclusively from the nuclear material (Figure 4.1D). Coomassie brilliant blue staining of the polyvinylidene difluoride (PVDF) membrane showed the equality in protein loading and the efficiency of transfer onto PVDF membrane across the protein gel (Figure 4.1E).

Based on these results, the subsequent proteomics experiment exploring the SUMO2 signalling in response to heat-activated proteotoxic stress was performed with HEK293$^{6His\text{-}SUMO2\text{-}T90K}$ cells incubated at 43 °C for 30 minutes.

4.2.2. Overview of the experimental strategy

To investigate the alterations in global protein sumoylation in response to heat-induced proteotoxic stress, a SILAC quantitative MS-based approach was undertaken (Figure 4.2) (Ong et al., 2002). For this purpose, HEK293$^{6His\text{-}SUMO2\text{-}T90K}$ cells were cultured in a growth medium supplemented with natural L-lysine ($^{12}C_6$, $^{14}N_2$; K0) and L-arginine ($^{12}C_6$, $^{14}N_4$; R0) (Figure 4.2; grey). The K0R0-labelled culture was divided into two equal portions and one of the subcultures was subjected to hyperthermic stress by elevating the incubation temperature to 43 °C for 30 minutes (Figure 4.2; orange). After harvesting, both subcultures were disrupted in a buffer containing 100 mM 2-chloroacetamide, and separated into nuclear and cytoplasmic fractions. Nuclei were collected by centrifugation and washed to reduce the amount of cytoplasmic compartment-derived "contaminant" proteins. Finally, the subcellular compartments

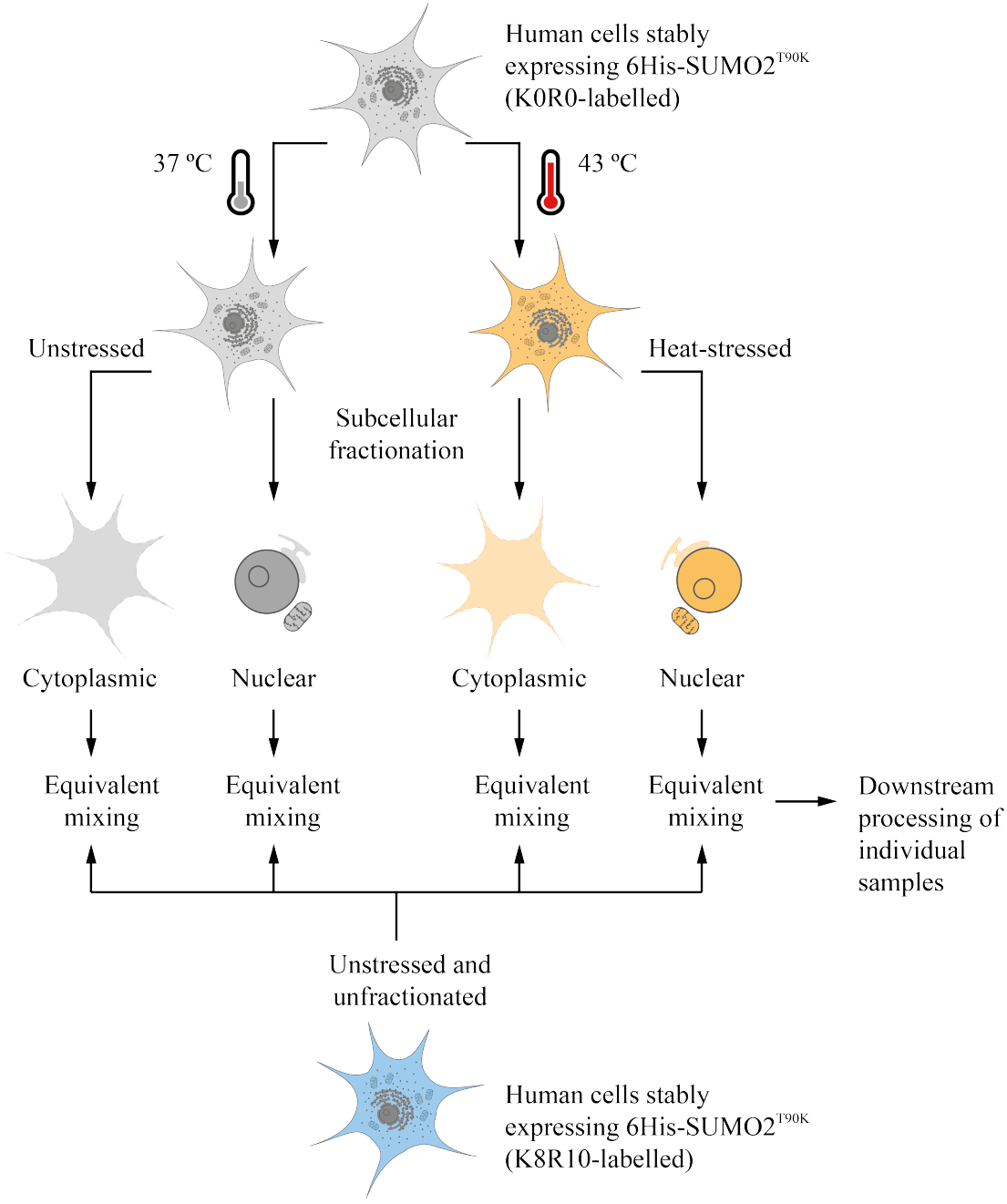

Figure 4.2 | Experimental strategy.

HEK293$^{6His\text{-}SUMO2\text{-}T90K}$ cells cultured in natural L-Lys and L-Arg-containing growth medium (K0R0; grey) were divided in two, and one of the subcultures was exposed to heat stress (43 °C, 30 min; orange). Both subcultures were then harvested and maintained separately during the subcellular fractionation procedure (top-down flowchart). In parallel, HEK293$^{6His\text{-}SUMO2\text{-}T90K}$ cells were cultured in heavy stable isotope-containing growth medium at 37 °C (K8R10; blue) and lysed directly without fractionation (bottom-up flowchart). An equivalent amount of K8R10-labelled untreated unfractionated cell lysate protein was then mixed with each K0R0-labelled sample and downstream processing was performed according to the SUMO^{KGG} workflow.

and protein structures in each analytical fraction were disrupted by the addition of a strong chaotropic agent Gu-HCl, and the concentration of proteins was determined using a bicinchoninic acid (BCA) assay (Figure 4.2; top-down flowchart).

In parallel, another culture of HEK293$^{6His\text{-}SUMO2\text{-}T90K}$ cells was metabolically labelled with heavy stable isotope-containing L-lysine ($^{13}C_6$, $^{15}N_2$; K8) and L-arginine ($^{13}C_6$, $^{15}N_4$; R10), and cultured at 37 °C (Figure 4.2; blue). The K8R10-labelled cells were harvested under strong denaturing conditions without fractionation and the protein concentration was determined using BCA assay (Figure 4.2; bottom-up flowchart). This sample of heavy-labelled proteins extracted from unstressed unfractionated cells was to serve as an internal standard that ultimately allowed direct comparison of individual protein abundances between all four fractions (subsequently termed analytes).

The internal standard was then spiked into each of the analytes according to a specific mixing strategy, which was designed to prevent systematic errors that could arise from the inequality in the initial number of cells cultured either at 37 °C or 43 °C (Equation 4.1). The set-up was devised to understand the reproducibility of the fractionation process, and to enable direct quantitative comparisons of individual protein abundances between all four analytes. Therefore, each fractionated sample was spiked with the amount of K8R10-labelled standard that corresponded to a half of the absolute protein extracted from either unperturbed or heat-stressed HEK293$^{6His\text{-}SUMO2\text{-}T90K}$ cells, respectively (Equation 4.1).

Equation 4.1 | Strategy for the spiking of standard protein.

$$H_S = (L_N + L_C)/2$$

H_S, K8R10-labelled standard protein to be mixed with each analyte (mg)

L_N, nuclear analyte of a K0R0-labelled cell culture (mg)

L_C, cytoplasmic analyte of the same K0R0-labelled cell culture (mg)

Next, all generated mixtures of proteins were examined twice essentially as illustrated in (Figure 3.19). The first analysis was aimed to describe the changes in protein abundance or localisation within human cells after the exposure to elevated temperature. Thus, a small proportion of each sample was subjected directly to in solution digestion with Lys-C, either alone or in tandem with trypsin. The emerging mixtures of peptides were fractionated offline by strong anion exchange (SAX) chromatography and analysed by LC-MSMS. Finally, four analyte-to-standard SILAC ratios were acquired by comparing the relative abundances of individual proteins between the K8R10-labelled standard and the K0R0-labelled cytoplasmic or nuclear fraction of cells cultured either at 37 °C or 43 °C. Ultimately, these SILAC ratios represented the subcellular distribution of the human proteome in normal or heat-stressed HEK293$^{6His\text{-}SUMO2\text{-}T90K}$ cells. These ratios were also used to confirm the equality in the amount of internal standard spiked into each analyte and when required, to reduce the systematic errors by normalisation.

The remaining mixtures of proteins were processed according to the SUMO^{KGG} workflow developed in Chapter 3. Analogously to the abovementioned samples, four analyte-to-standard SILAC ratios were acquired. However, these were designed to provide site-specific information about the global reprogramming of protein

sumoylation in response to heat-activated proteotoxic stress. Importantly, the entire workflow was performed in duplicate to improve data reliability.

4.2.3. Preliminary examination of the quality of fractionated protein samples

Prior to mixing, all analytes were subjected to the Western blot analyses for the detection of SUMO2/3, 6His, Tubulin, or Lamin A/C to ensure the efficiency of the heat shock and the subcellular fractionation procedure (Figure 4.3). As anticipated, the sumoylated substrates were predominantly detected from the high-molecular-weight area of the nuclear fraction, whereas the unconjugated form of SUMO2/3 was mainly in the cytoplasmic material (Figure 4.3A and B; cytoplasmic; nuclear). Also, an extensive heat-induced increase in substrate-bound SUMO2/3 was observed in the nuclear analyte. The separation of sumoylated proteins and unconjugated SUMO2/3 between the subcellular fractions was remarkably clear-cut, and only low levels of monomeric SUMO2/3 were detected in the wash (Figure 4.3A and B; wash). Heavy-labelled proteins from unstressed and unfractionated cells contained both, unconjugated and substrate-bound SUMO2/3 (Figure 4.3A and B; unfractionated).

The distribution of Lamin A/C or Tubulin verified the efficient fractionation (Figure 4.3C and D). As expected, Tubulin was mainly cytoplasmic, whereas Lamin A and Lamin C were predominantly nuclear. Out of the four immunoblot images, only Tubulin was consistently detected from the wash fraction, and illustrated the importance of washing the extracted nuclei to eliminate potential contamination (Figure 4.3C; wash). However, less than 7.5 % of the protein extracted per cell culture was on average measured from the wash fraction.

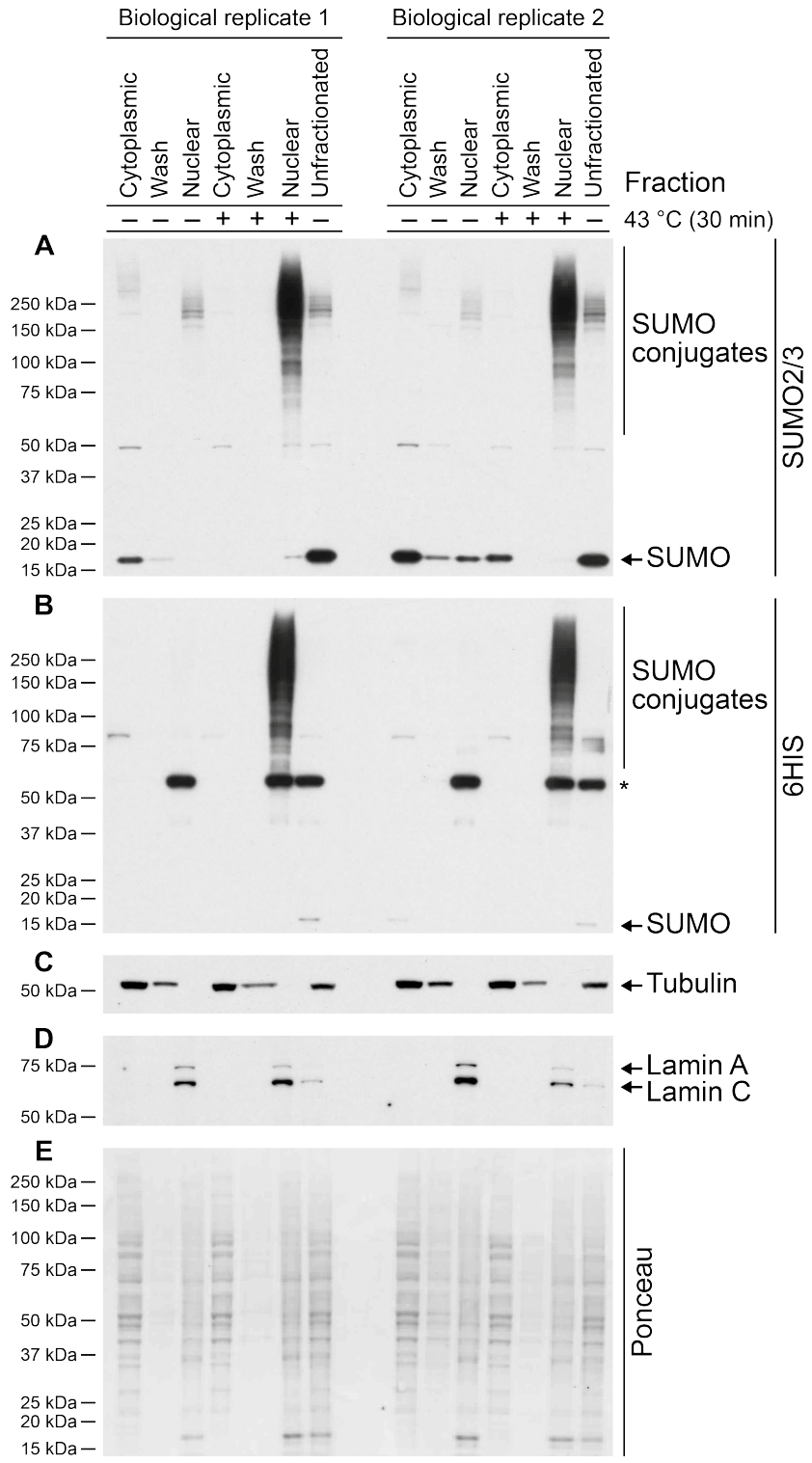

Figure 4.3 | Increased SUMO2 conjugation in the nuclear fraction of heat-stressed cells.

Fractions of cellular proteins generated according to the experimental strategy presented in Figure 4.2 were subjected to Western blot analyses for the detection of: (**A**) SUMO2/3, (**B**) 6His, (**C**) Tubulin, and (**D**) Lamin A/C. (**E**) Ponceau S staining of a PVDF membrane validating the equal loading and consistent transfer of proteins.

The detection of both Tubulin and Lamin C in the heavy-labelled material confirmed the unfractionated nature of the internal standard (Figure 4.3C and D; unfractionated). The Ponceau-stained PVDF membrane verified the equal loading and efficiency of transfer (Figure 4.3E).

The analytes were then mixed with the heavy-labelled standard and the experiment was conducted according to the devised protocol (Figure 4.2). Ultimately, four SILAC ratios reflecting the subcellular distribution of either 6His-SUMO2^{T90K}-modified or overall human proteome, were obtained in duplicate:

- "Cytoplasmic fr. at 37 °C"-to-"Internal standard"
- "Nuclear fr. at 37 °C"-to-"Internal standard"
- "Cytoplasmic fr. at 43 °C"-to-"Internal standard"
- "Nuclear fr. at 43 °C"-to-"Internal standard"

4.2.4. A comprehensive validation of the sample preparation workflow

As described in Chapter 4.2.2, all four analytes generated in duplicate were spiked with an unfractionated internal standard to enable direct comparison of individual protein abundances between each subcellular fraction of normal or heat-stressed human cells. However, a specific mixing scenario had to be devised to compensate for the potential inequalities in the number of cells utilised per experimental condition (Equation 4.1). Consequently, various amounts of internal standard were combined with analytes extracted from distinct cell cultures. The accuracy and reproducibility of

this mixing strategy had thus to be evaluated to assure its completion without systematic errors, or if required, to assist with normalisation.

The accuracy and precision of the spiking procedure were thus assessed by comparing the quantities of individual proteins in the analytical or standard sample according to Equation 4.2. Given that no systematic errors were introduced and the absolute abundances of most proteins remained unaltered in response to hyperthermia, the median value of these calculated relative abundances should be equal between cultures incubated at 37 °C or 43 °C. On average, this median value varied by merely ~2.9 %, illustrating the high precision of the spiking procedure (Table 4.1; coefficient of variation). However, small imperfections in the accuracy were observed, as the geometric averages of the median relative abundances within duplicate experiments were either 1.3 (± 0.5 %) or 0.7 (± 5.4 %), instead of the anticipated value of one (Table 4.1; median of relative abundances). The analyte-to-standard ratios of all proteins were thus normalised to diminish the introduced spiking-related systematic errors.

Equation 4.2 | Strategy for the assessment of the spiking accuracy.

$$(L_N + L_C)/(2 \times H_S) = 1$$

H_S, K8R10-labelled standard protein mixed with each analyte (mg)

L_N, nuclear analyte of a K0R0-labelled cell culture (mg)

L_C, cytoplasmic analyte of the same K0R0-labelled cell culture (mg)

Table 4.1 | Precision and accuracy of spiking procedure.

Statistic	First study		Second study	
Culturing temperature (°C)	37	43	37	43
Median of relative abundances[a]	1.3	1.3	0.7	0.7
Coefficient of variation (%)[b]	0.5		5.4	

[a] Calculated according to Equation 4.2; the measurement is based on 50 % of proteins with highest MS signal intensity

[b] Measured between the cultures within each study

The reproducibility of the fractionation procedure was evaluated using the four analyte-to-standard SILAC ratios acquired in duplicate. Consistency in the distribution of individual proteins between two subcellular fractions was measured by determing the cell culture-specific “Nuclear fr.”-to-“Cytoplasmic fr.” relative protein abundances according to (Equation 4.3). This was compared between the duplicate experiments performed with cells either cultured at 37 °C or heat-stressed at 43 °C (Figure 4.4). As with all SILAC ratios, the computed “Nuclear fr.”-to-“Cytoplasmic fr.” ratios followed a lognormal distribution and were characterised by a positive skew. Therefore, each calculated ratio was log-transformed (base 2) to follow a normal distribution, which ultimately supported the straightforward visualisation of data and interpretation of relationships. The same approach was undertaken throughout this study and protein ratios were reported as log-transformed (base 2) values.

Equation 4.3 | Calculation of culture-specific "Nuclear fr."-to-"Cytoplasmic fr." protein ratios.

$$\frac{\text{Nuclear fr.}}{\text{Cytoplasmic fr.}} = \frac{\text{Nuclear fr.}}{\text{Internal standard}} \div \frac{\text{Cytoplasmic fr.}}{\text{Internal Standard}}$$

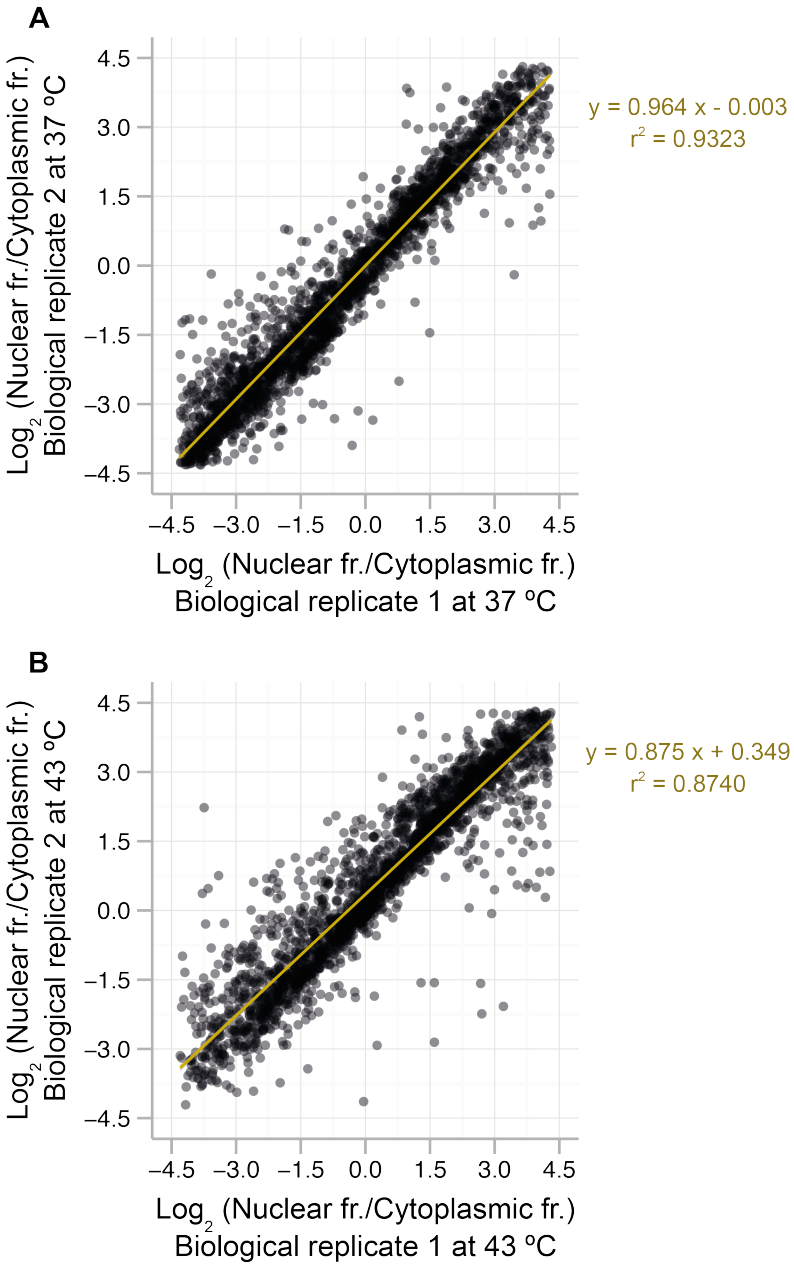

Figure 4.4 | Reproducibility of subcellular fractionation.

The consistency in the segregation of proteins into subcellular fractions was compared between the duplicate experiments performed with either unperturbed (**A**; 37 °C) or heat-stressed (**B**; 43 °C) HEK293$^{6His\text{-}SUMO2\text{-}T90K}$ cells. Both scatter plots were fitted with a least squares linear regression curve.

Two scatter plots were generated by comparing the log-transformed relative subcellular abundances of individual proteins between the duplicate experiments (Figure 4.4). Best-fit linear regression curves were modelled and R squared statistics

(r^2) were computed to assess the goodness of the fitted models. Considering that SILAC is particularly suitable for measuring relatively small protein changes and the less precise extreme outliers can have major impact on the linear regression line, only proteins with less than 20-fold relative difference between two subcellular fractions were utilised to evaluate the reproducibility of the fractionation procedure (Bantscheff et al., 2007).

In general, both linear regression lines reflecting the relationships between duplicate experiments were characterised by high R squared (r^2) values (Figure 4.4). Therefore, most of the observed variances between the duplicate experiments were explained by the modelled linear regression curves, which thus demonstrated the high consistency of the fractionation technique. Furthermore, the infinitesimal intercept value of the regression line modelled based on the subcellular distribution of proteins in cells cultured at 37 °C (Figure 4.4A) illustrated the precision in the amplitude of protein segregation achieved. In contrast, the systematic variance between duplicate fractionation procedures performed with heat-stressed cells was ~27.4 % (Figure 4.4B). Whether this variance was introduced by disproportionate fractionation efficiency or marginally different magnitude of hyperthermic-stress response cannot be resolved using the available information. Thus, elimination of this systematic variance by normalisation was not justified. Furthermore, it was speculated that the less than 1.3-fold systematic variance between duplicate experiments would not impair the general accuracy of quantification.

Finally, the performance of the workflow to represent the genuine snapshot of the spatial distribution of the human proteome was examined. To this end, the

immunofluorescence microscopy (IF)-based subcellular annotation data was retrieved from the Human Protein Atlas database (HPA), which is a rich resource of complementary localisation information collected with an alternative technology (Uhlen et al., 2015). Every protein in the HPA database has been classified based on the reliability of the annotation. Only "Supportive" entries, whose immunostaining pattern has been validated either by multiple antibodies or by the protein characterisation information available in the UniProt database, were used for the comparisons (The UniProt Consortium, 2015). Finally, the concurrency of the annotations assigned by either immunofluorescence microscopy or mass spectrometry was evaluated.

Considering the continuous nature of MS-based data, a frequency histogram was generated to reflect the subcellular distribution of all proteins quantified in this study (Figure 4.5). In particular, the quantified proteome of HEK293$^{6His\text{-}SUMO2\text{-}T90K}$ cells was partitioned into 40 bins based on the individual "Nuclear fr."-to-"Cytoplasmic fr." ratios, and the number of proteins in each bin was counted. The histogram bins were then coloured according to the porportion of proteins with a primary IF-based annotation as cytoplasmic (green), mitochondrial (blue), nuclear (maroon), or multiple locations (light grey) (Figure 4.5).

As illustrated by the stacked histogram (Figure 4.5), the overall MS-based subcellular allocation of the proteome of HEK293$^{6His\text{-}SUMO2\text{-}T90K}$ cells followed a bimodal distribution, suggesting that most proteins have a distinct primary location either in the cytoplasmic or nuclear compartment. On the other hand, less than 1.5-fold difference between the compartment-specific abundances was observed for 15.9 % of the quantified proteins, indicating either a dynamic location or function in multiple subcellular loci.

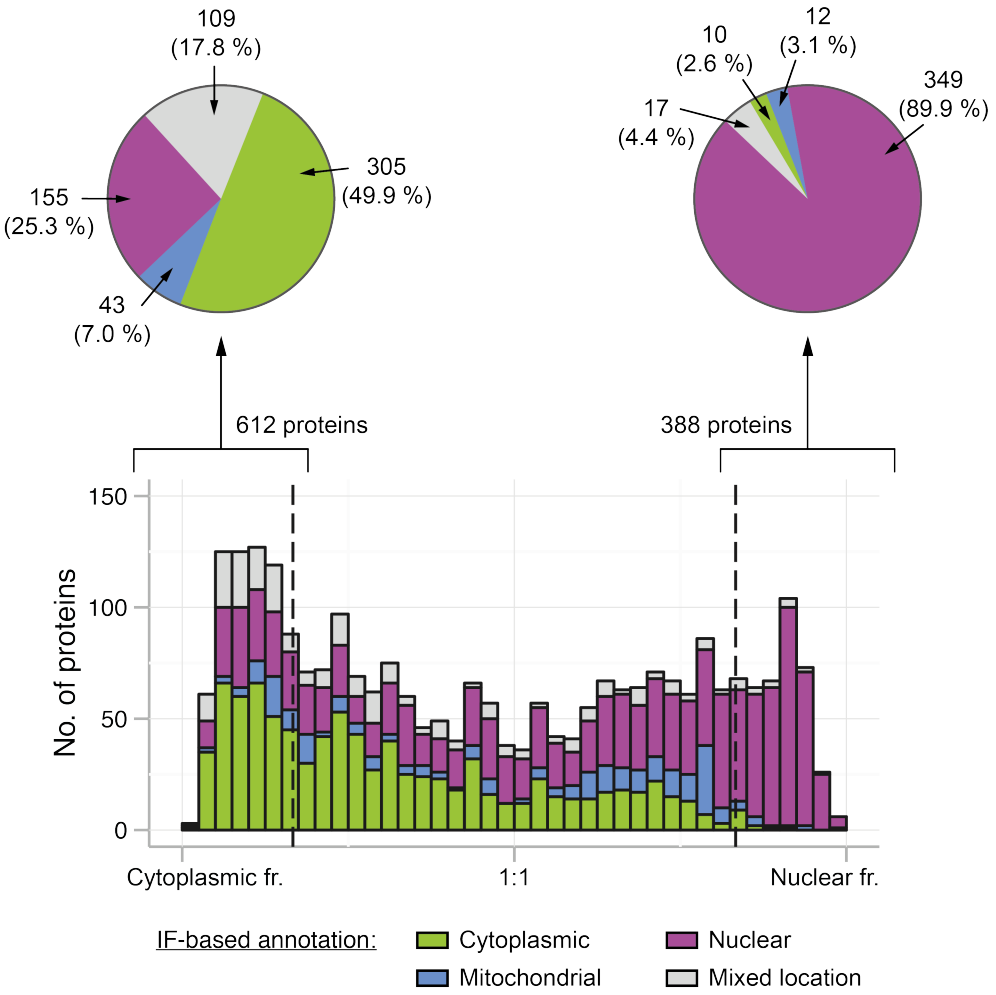

Figure 4.5 | Efficiency of subcellular fractionation.

Stacked frequency histogram of the subcellular distribution of HEK293$^{6His\text{-}SUMO2\text{-}T90K}$ proteome at 37 °C. Coloured areas represent the proportion of proteins with cytoplasmic (green), mitochondrial (blue), nuclear (maroon), or unambiguous (grey) immunofluorescence microscopy (IF)-based location annotation retrieved from Human Protein Atlas database (Uhlen et al., 2015). Two pie charts were created to analyse the distribution of each IF-based annotation among proteins, whose relative abundances between the compartments differed by at least fivefold.

In general, the IF-based annotations were in agreement with the MS-based spatial allocation. As expected, proteins with mainly cytoplasmic IF pattern had a tendency to segregate into the cytoplasmic subcellular fraction (Figure 4.5; green), and no significant sub-compartment- or organelle-specific variances were detected (Figure

4.6A–C). Surprisingly, the subcellular localisation of proteins with a mitochondrial IF pattern was similar to the global histogram and was characterised by a bimodal distribution (Figure 4.5 and Figure 4.6D; blue). However, multiple independent research groups have claimed that the phenomenon is a fractionation technique-specific artifact and the separation of mitochondria from nuclei is inefficient (Boisvert et al., 2012; Raychaudhuri et al., 2014). Nevertheless, ~89.9 % of the 388 proteins with at least 5-fold greater abundance in the nuclear fraction were in agreement with their IF-based annotation, thus demonstrating the high purity of the fraction of nuclear proteins (Figure 4.5; second pie chart).

In contrast, a significant amount of proteins with mainly nuclear IF signal were more abundant in the cytoplasmic subcellular fraction of the MS-based study (Figure 4.5 and Figure 4.6E–G; maroon). In particular, nuclear IF annotation was retrieved for 25.3 % of the 612 proteins, whose MS-based abundance was more than 5-fold greater in the cytoplasmic fraction (Figure 4.5; first pie chart). A detailed inspection revealed that many of those 155 proteins, such as the 12 subunits of the proteasome or six components of the Constitutive photomorphogenesis 9 (COP9) signalosome, are present in both compartments. Moreover, their subcellular stoichiometry is often cell line-specific (Rivett et al., 1992; Palmer et al., 1996; Fuzesi-Levi et al., 2014). However, this type of annotation discrepancy was observed for many other proteins and the reasons for those are still unclear. Therefore, caution should be applied when interpretating data, which is based on proteins with ambiguous subcellular location.

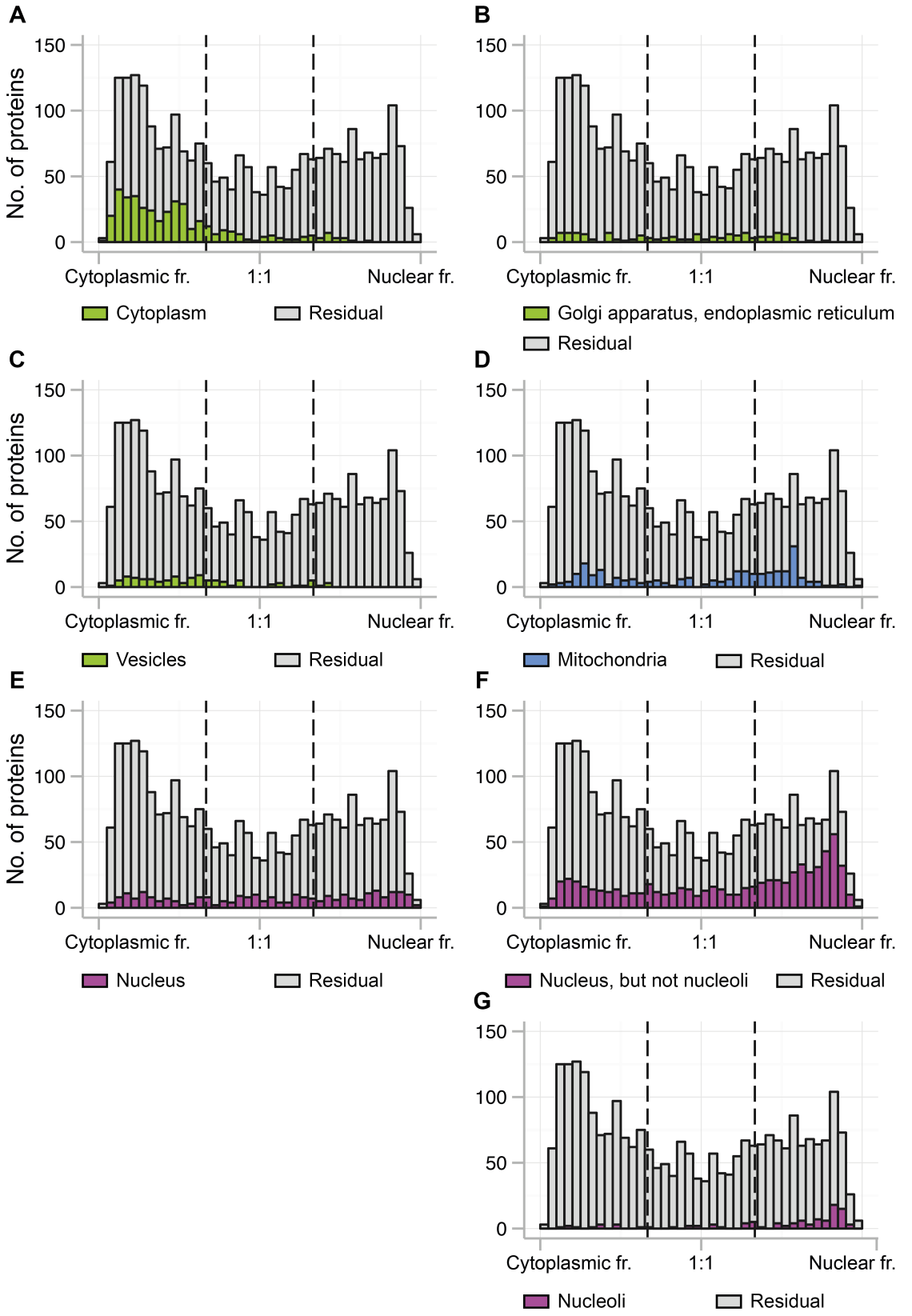

Figure 4.6 | Spatial distribution of the HEK293$^{6His\text{-}SUMO2\text{-}T90K}$ proteome grouped by the IF-based location annotation.

››

‹‹ Figure 4.6 | Spatial distribution of the HEK293$^{6His\text{-}SUMO2\text{-}T90K}$ proteome grouped by the IF-based location annotation.

Frequency histograms of the subcellular distribution of the HEK293$^{6His\text{-}SUMO2\text{-}T90K}$ proteome coloured according to the IF-based location retrieved from the Human Protein Atlas database (Uhlen et al., 2015). Coloured area of each distribution represents the proportion of proteins that according to the HPA annotation are predominantly located in: (**A**) cytoplasm, (**B**) golgi apparatus or endoplasmic reticulum, (**C**) vesicles, (**D**) mitochondria, (**E**) nucleus, (**F**) nucleus, but are excluded from nucleoli, (**G**) nucleoli. The colour code is linked to Figure 4.5 with green, blue and maroon representing subcellular compartments that were combined into cytoplasmic, mitochondrial or nuclear IF-based annotation groups, respectively.

4.2.5. Uncovering the dynamics of the human proteome in response to heat-induced proteotoxic stress

To reveal the acute heat shock-induced global changes to the spatial distribution of the human proteome, a small proportion of each SILAC-labelled mixture of proteins (Figure 4.2) was processed and analysed by LC-MSMS as described in Chapter 4.2.2. In total, 8024 proteins were identified and 63.5 % of them were quantified in every analyte. Using the acquired SILAC ratios, "Nuclear fr."-to-"Cytoplasmic fr." values were calculated (Equation 4.3) and the subcellular distribution was determined for 5099 proteins expressed in HEK293$^{6His\text{-}SUMO2\text{-}T90K}$ cells (Supplementary table S2; Appendix 6.1). Furthermore, as each of these proteins was measured in all analytes, potential changes to their location or absolute abundance in cells exposed to heat-induced proteotoxic stress could be assessed.

First, the fluctuations in the absolute amount of individual proteins in response to hyperthermia were examined. To this end, the internal standard-normalised subcellular fraction-specific abundances were added and the summed values were

compared between samples extracted from either normal or heat-stressed HEK293$^{6His\text{-}SUMO2\text{-}T90K}$ cells. After a 30-minute treatment, the absolute quantity of nearly all proteins remained unaltered (Figure 4.7; Table 4.2). Notably, the observation is consistent with the study based on TAP-SUMO2-expressing HeLa (human cervix epitheloid carcinoma) cells, which were cultured at 43 °C for an equal period of time (Golebiowski et al., 2009).

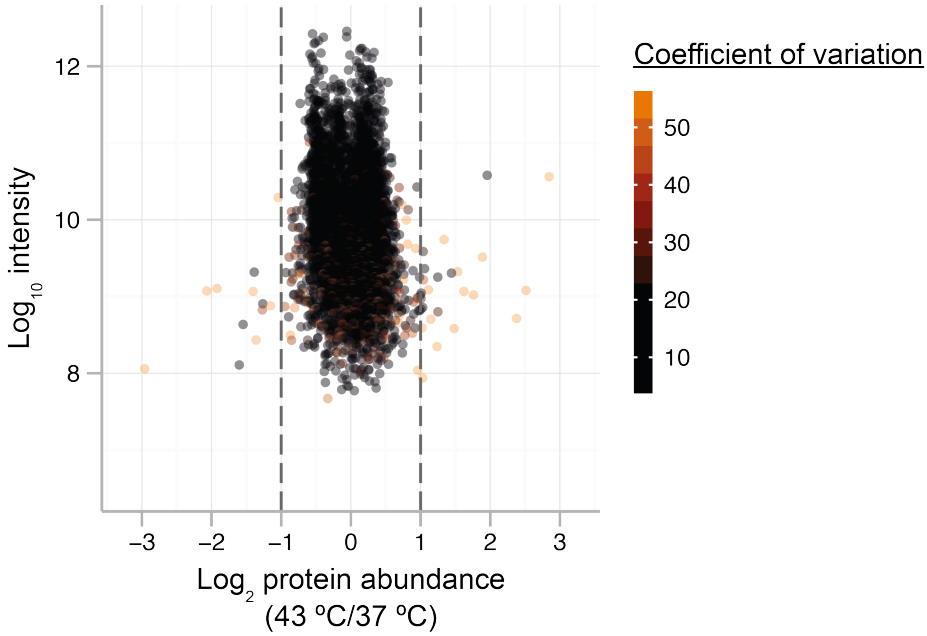

Figure 4.7 | The abundance of proteins remains unaltered in cells exposed to heat stress.
Scatter plot of log-transformed (base 2) relative protein abundances in heat-stressed or unperturbed HEK293$^{6His\text{-}SUMO2\text{-}T90K}$ cells (43 °C/37 °C) with y-axis representing the log-transformed (base 10) summed MS signal intensity. Each data point corresponds to a single protein and is displayed using a colour gradient from black to orange according to the variability of its relative ratio in replicate experiments in an increasing order.

Table 4.2 | Proteins with greater than twofold change in absolute abundance.

Seq. no.	Protein name	Gene name	Log_2 protein abundance (43 °C/37 °C)[a]	Coefficient of variation (%)[b]
1	Nucleoside diphosphate kinase, mitochondrial	NME4	1.44350	3.68
2	Inner centromere protein	INCENP	1.33760	57.81
3	Coiled-coil domain-containing protein 14	CCDC14	1.24918	26.53
4	HAUS augmin-like complex subunit 2	HAUS2	1.24670	13.76
5	Centromere protein J	CENPJ	1.11334	56.22
6	Protein KHNYN	KHNYN	1.05692	4.21
7	Sugar phosphate exchanger 3	SLC37A3	1.04173	42.93
8	Short-chain specific acyl-CoA dehydrogenase, mitochondrial	ACADS	1.03793	22.32
9	tRNA-dihydrouridine(16/17) synthase [NAD(P)(+)]-like	DUS1L	1.03051	12.19
10	Cytochrome b-c1 complex subunit 8	UQCRQ	1.00845	15.09
11	Protein CASC5	CASC5	1.00599	21.28
12	Junction-mediating and regulatory protein	JMY	1.00566	26.88
13	RILP-like protein 1	RILPL1	-1.26435	18.99
14	Geranylgeranyl transferase type-1 subunit beta	PGGT1B	-1.27096	39.82
15	E3 ubiquitin-protein ligase MARCH6	MARCH6	-1.38700	16.30
16	Arf-GAP with SH3 domain, ANK repeat and PH domain-containing protein 2	ASAP2	-1.54726	19.85
17	Histone-lysine N-methyltransferase ASH1L	ASH1L	-1.60220	9.55

[a] Each protein quantification is based on at least three peptides

[b] Includes proteins with coefficient of variation up to 59.35 %

To determine the subcellular compartment-specific abundances of individual proteins, two complementary methods were employed. First, the internal standard-normalised abundances of proteins either in the nuclear or cytoplasmic fraction of normal cells were compared with their heat-stressed counterparts. Thus, the analysis was designed to reveal the temperature-induced subcellular compartment-specific fold changes to individual protein abundances (Figure 4.8 and Figure 4.9; abscissa). Secondly, the "Nuclear fr."-to-"Cytoplasmic fr." ratios were utilised to define the proportion of the total pool of each protein in the two subcellular compartments in normal cells. This distribution was then compared to the heat-stressed cells, and thus the proportion of each protein pool, whose location was affected by the elevated temperature, was uncovered (Figure 4.8 and Figure 4.9; ordinate).

Surprisingly, a substantial movement of proteins from cytoplasmic to nuclear fraction was observed (Figure 4.8 and Figure 4.9). In particular, more than twofold increase in the nuclear analyte was reproducibly measured for 1129 proteins (Figure 4.8; abscissa). As expected, all SUMO paralogs shuttled in response to hyperthermic stress, as the nuclear abundance of SUMO1 and SUMO2/3 increased by approximately 3.3- or 4.4-fold, respectively (Figure 4.8; Supplementary table S2; Appendix 6.1). In eukaryotic cells, the activation of heat shock response is mediated by a transcription factor HSF1 (Heat shock factor protein 1) that regulates the expression of a subset of genes encoding molecular chaperones (Hartl et al., 2011; Solis et al., 2016). In normal cells, HSF1 is held as an inactive monomer in the cytoplasm and nucleus. However, upon cellular stress the monomeric molecules of HSF1 are converted into high-affinity DNA-binding homotrimers, which translocate to nucleus and activate the transcription of

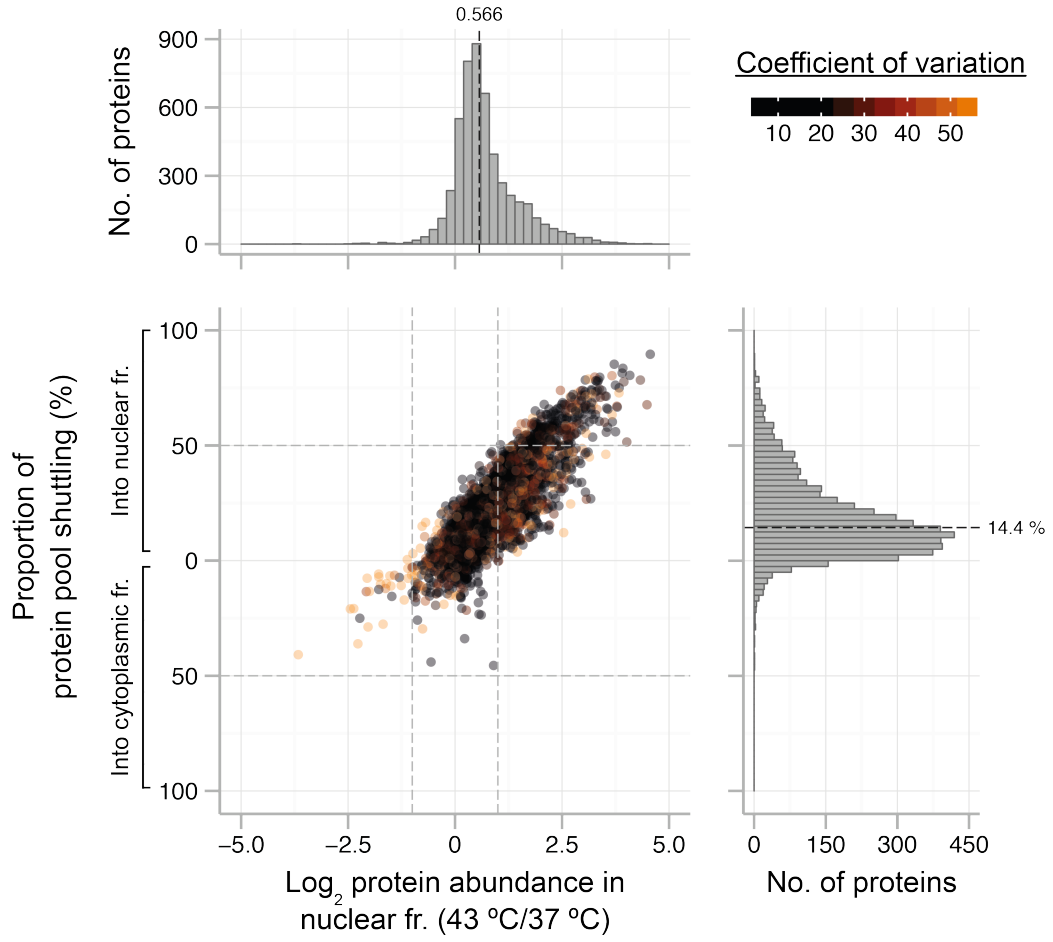

Figure 4.8 | The abundance of 1129 proteins is increased in the nuclear fraction by more than twofold in response to heat shock.

Scatter plot of log-transformed (base 2) relative protein abundance in the nuclear subcellular fraction of heat-stressed or unperturbed HEK293$^{6His\text{-}SUMO2\text{-}T90K}$ cells (x-axis; 43 °C/37 °C). Y-axis represents the percentage of the absolute pools of individual proteins that shuttle in response to acute heat shock. The distributions of x and y values are represented by frequency histograms. Each data point is displayed using a colour gradient from black to orange according to the increasing variability of replicate relative abundance values.

genes encoding heat shock proteins (HSPs) (Baler et al., 1993; Sarge et al., 1993; Akerfelt et al., 2010). In agreement with these characteristics, the nuclear abundance of HSF1 was increased by approximately 8.2-fold in HEK293$^{6His\text{-}SUMO2\text{-}T90K}$ cells cultured at 43 °C for 30 minutes (Figure 4.8; Supplementary table S2; Appendix 6.1).

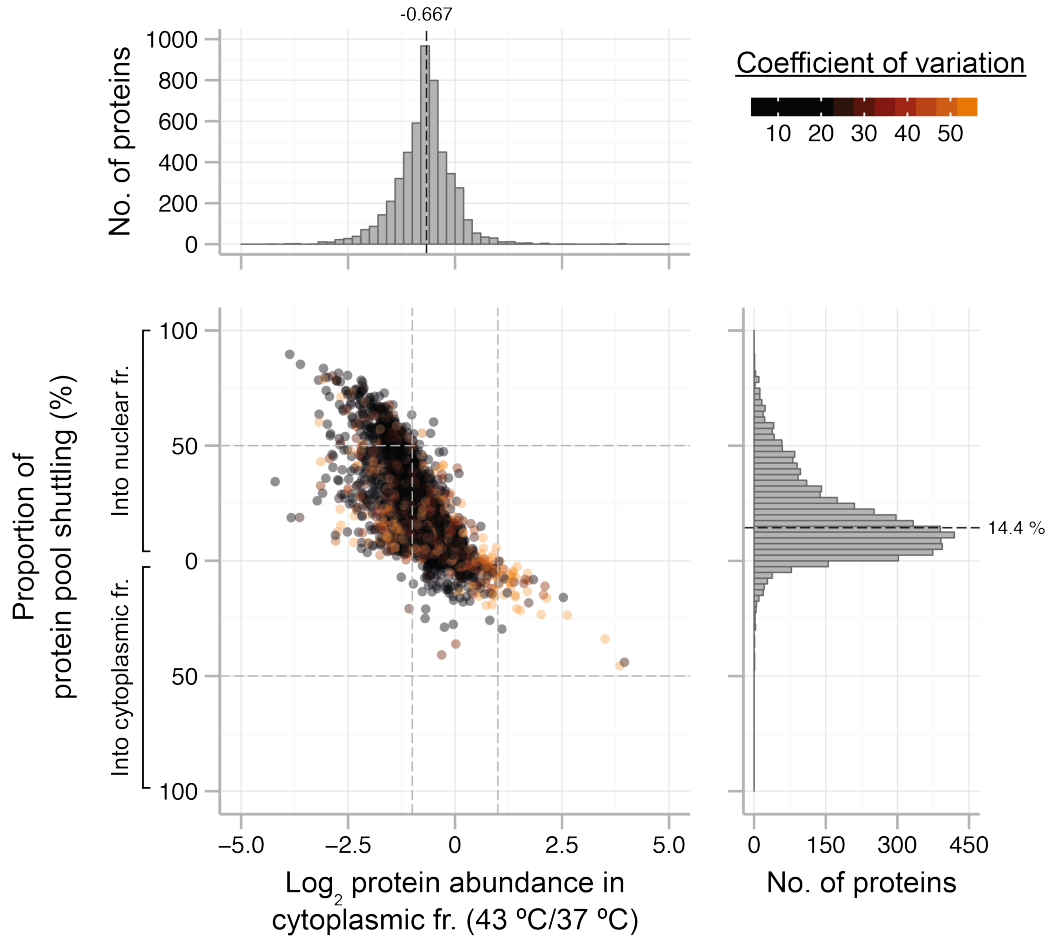

Figure 4.9 | The cytoplasmic abundance of 1231 proteins decreased by more than twofold in heat-stressed human cells.

Scatter plot of log-transformed (base 2) relative protein abundance in the cytoplasmic fraction of heat-shocked or normal HEK293$^{6His\text{-}SUMO2\text{-}T90K}$ cells (x-axis; 43 °C/37 °C). Y-axis illustrates the percentage of individual protein pools that shift in response to heat stress. The distributions of both data sets are represented by frequency histograms. Each data point corresponds to a single protein and is displayed using a colour gradient from black to orange according to the increasing variability of relative abundances between replicate experiments.

In contrast with the general increase of proteins in the nuclear subcellular fraction, an expected global reduction in the individual protein quantities was detected in the cytoplasmic subdivision of the heat-stressed cells (Figure 4.9). Specifically, the abundance of 1231 proteins was reproducibly reduced in the cytoplasmic fraction by more than twofold (Supplementary table S2; Appendix 6.1). An inverse relationship

was observed for 11.6 % of quantified proteins, which were simultaneously increased in the nuclear and decreased in the cytoplasmic fraction. This suggests that the proportion of proteins, which substantially translocate between the two subcellular fractions in response to acute heat stress is low. Indeed, the observation was further confirmed by the second analysis, which specifically measured the proportions of individual protein pools that shift between the cell fractions in response to heat shock (Figure 4.8 and Figure 4.9; ordinate). Although the subcellular compartment-specific abundances of many proteins were altered by more than twice, these fold changes were mostly created by minor proportions of the entire protein pools. However, at least 14.4 % of the total pool of 2550 quantified proteins was shifted into nuclear fraction in response to hyperthermia. Moreover, 245 proteins were substantially and reproducibly translocated, with at least 50 % of their entire pool moving into nuclear compartment (Figure 4.8 and Figure 4.9). However, within this group there was no obvious enrichment of proteins with distinct molecular activity, or an affiliation with specific biological pathways or macromolecular complexes (Ashburner et al., 2000; Ruepp et al., 2008; The Gene Ontology Consortium, 2015).

4.2.6. Spatial distribution of HSF1 and SUMO2

As discovered in Chapter 4.2.4, the assigned subcellular annotation of some proteins can be ambiguous. One of these proteins with inconclusive location was the critical regulator of heat stress response: transcription factor HSF1. As this project investigates the spatial distribution of the human proteome in response to hyperthermia, the bona fide location of HSF1 itself in normal and heat-stressed cells should be unambiguously assigned.

According to this study, nearly 85 % of the HSF1 molecules were segregated into the cytoplasmic fraction of HEK293$^{6His\text{-}SUMO2\text{-}T90K}$ cells cultured at 37 °C (Supplementary table S2; Appendix 6.1). Conversely, the immunofluorescence microscopy-based annotation retrieved from the Human Protein Atlas database suggests that HSF1 is mainly nuclear (Uhlen et al., 2015). To resolve this discrepancy, an immunofluorescence microscopy experiment was conducted using HeLa cells, which considering their size, morphology and better adherence are more suitable for the IF-based studies than the HEK293$^{6His\text{-}SUMO2\text{-}T90K}$ N3S cells.

Surprisingly, under normal conditions, the majority of HSF1 was detected in nuclei and was characterised by a diffused IF pattern (Figure 4.10; 37 °C). However, the immunostaining of HSF1 changed considerably in response to heat-induced proteotoxic stress, as large intense foci were formed (Figure 4.10; 43 °C). These foci, known as the nuclear stress bodies, have been described to appear transiently in stressed human cells and are considered to be the predominant location of HSF1 accumulation upon heat shock (Cotto et al., 1997). The generated images were thus in agreement with the IF-based annotation of the location of HSF1.

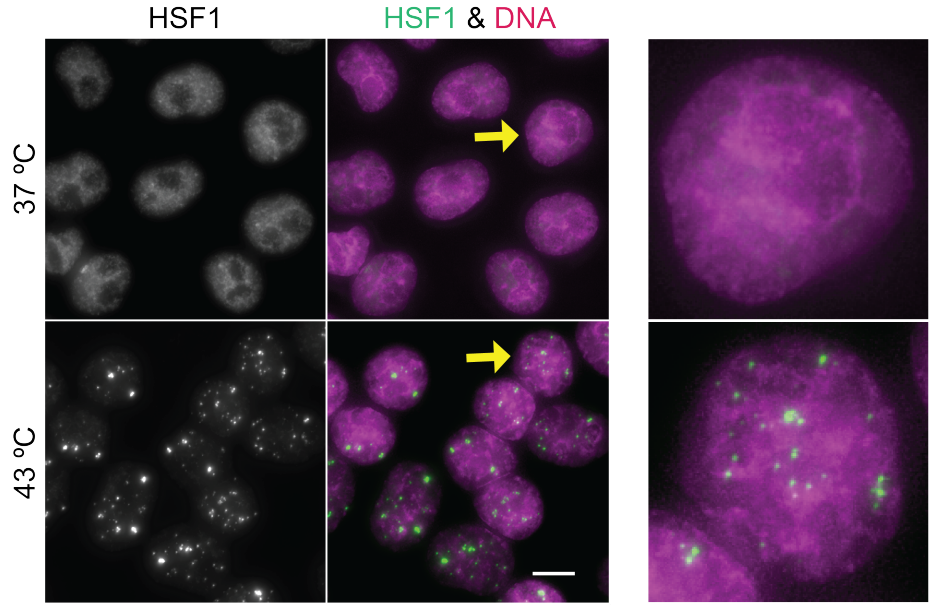

Figure 4.10 | HSF1 accumulates in subnuclear foci in response to heat stress.
Unperturbed (37 °C) or heat-stressed (43 °C) HeLa cells were analysed by immunofluorescence microscopy with HSF1-specific rabbit polyclonal antibody (green). DNA was stained with DAPI (maroon). Scale bar represents 10 μm.

According to the Western blot images of cellular fractions, the unconjugated molecules of SUMO2/3 were predominantly segregated into the cytoplasmic fraction of the HEK293$^{6His\text{-}SUMO2\text{-}T90K}$ cells (Figure 4.3A and B). Although the protein is not annotated in the Human Protein Atlas database, the IF-based information on SUMO2/3 localisation can be found from the literature, which suggests that the protein is mainly nuclear and accumulates in the subnuclear PML bodies (Vertegaal et al., 2004; Tatham et al., 2008; Uhlen et al., 2015). Another immunofluorescence microscopy experiment was thus performed to resolve this apparent discrepancy and pinpoint the genuine location of SUMO2/3 in normal and heat-stressed human cells.

Consistent with the previous IF-based studies, predominantly nuclear SUMO2/3-specific signal was detected under normal conditions (Figure 4.11; 37 °C). Moreover,

the recruitment of SUMO2/3 to subnuclear bodies was illustrated by the existence of SUMO2/3-containing foci. However, similarly to the HSF1, the image was also characterised by a diffused IF pattern. Interestingly, most of the diffuse SUMO2/3 was decreased and many small subnuclear foci were formed after the activation of heat shock response (Figure 4.11; 43 °C).

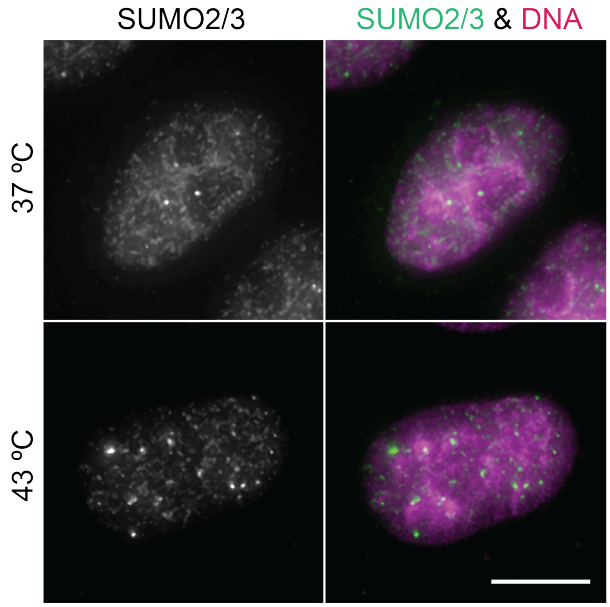

Figure 4.11 | SUMO2/3 accumulates in subnuclear foci in response to heat stress.
The localisation of SUMO2/3 in unperturbed (37 °C) or heat-stressed (43 °C) human cells assessed by immunofluorescence microscopy using a rabbit monoclonal antibody (green). DNA was stained with DAPI (maroon). Scale bar represents 10 μm.

The produced immunofluorescence microscopy images thus suggest that HSF1 and SUMO2/3 are predominantly nuclear irrespective of the culturing temperature of the human cells. Moreover, both proteins accumulate in subnuclear foci in response to hyperthermic stress. Interestingly, this recruitment into subnuclear foci correlates with the increased amount of SUMO2/3 and HSF1 measured in the nuclear fraction of HEK293$^{6His\text{-}SUMO2\text{-}T90K}$ cells (Supplementary table S2; Appendix 6.1).

4.2.7. Subcellular localisation and functional characterisation of targets of SUMO2

There is a consensus that sumoylation occurs predominantly in the nucleus (Kamitani et al., 1997; Rodriguez et al., 2001). However, a comprehensive high-throughput comparison between the spatial distributions of sumoylated or unmodified molecules of individual proteins has not been conducted. Furthermore, the equilibrium of protein localisation, either SUMO-conjugated or unmodified, could be affected in response to stress-inducing stimuli, such as the heat shock.

To quantitatively compare the spatial distributions of sumoylated proteins between unperturbed and heat-stressed human cells, the analytical samples generated in Chapter 4.2.2 (Figure 4.2) were processed according to the $SUMO^{KGG}$ workflow developed in Chapter 3 (Figure 3.8). The SUMO remnant-modified Gly-Gly-Lys-containing peptides were then analysed by LC-MSMS with settings adjusted for the optimal detection, identification and quantification of low abundance sumoylated sequences.

In total, 612 SUMO2 modification sites from 383 proteins were identified from $HEK293^{6His\text{-}SUMO2\text{-}T90K}$ cells cultured at 37 °C (Supplementary table S3; Appendix 6.1). As anticipated, approximately 88.5 % of the sumoylated lysines were detected solely from the nuclear fraction (Figure 4.12; abscissa). This is consistent with the Western blot analyses described in Chapter 4.2.3 (Figure 4.3). Functional characterisation of these 383 modified proteins revealed that nearly 40.0 % of them are involved in transcription of DNA. Furthermore, SUMO modification capability was identified for several

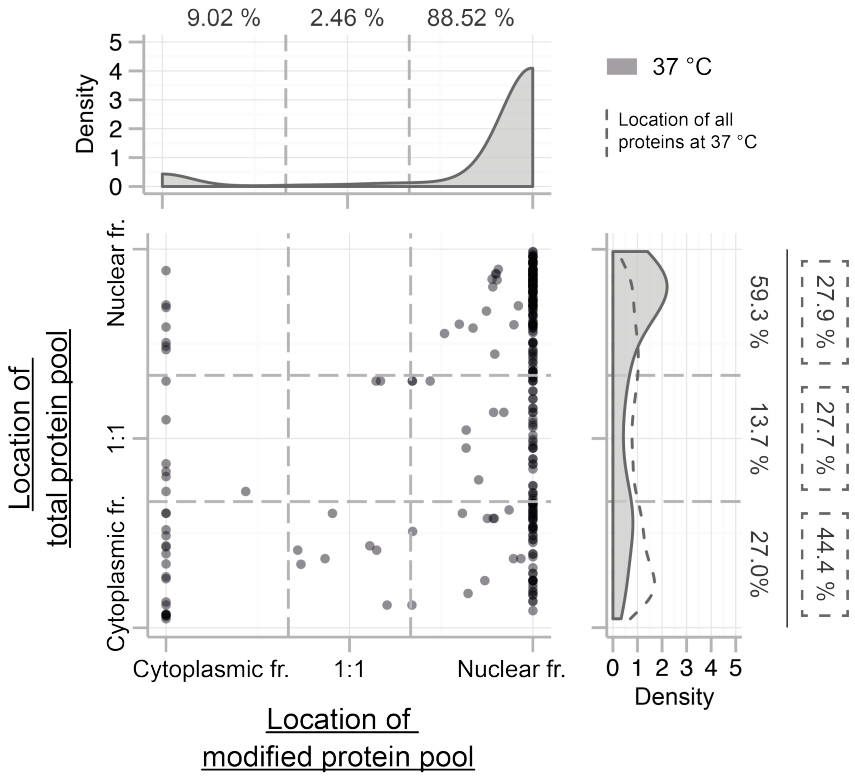

Figure 4.12 | 6His-SUMO2^{T90K} conjugates localise to the nuclear fraction of unstressed cells. Scatter plot with subcellular localisations of individual sumoylated proteins (x-axis) compared to the spatial distributions of the total pools of corresponding proteins (y-axis) in HEK293$^{6His\text{-}SUMO2\text{-}T90K}$ cells cultured at 37 °C. The kernel density plots display the probability density function of the subcellular distribution of either SUMO-modified (top; filled area) or total (right; filled area) pool of the sumoylated proteome. The probability density function of the localisation of all 5099 proteins quantified in cells cultured at 37 °C is illustrated by the dashed line (right).

proteins linked with fundamental cellular mechanisms, such as cell cycle progression, RNA processing, or DNA replication, recombination and repair (Figure 4.13).

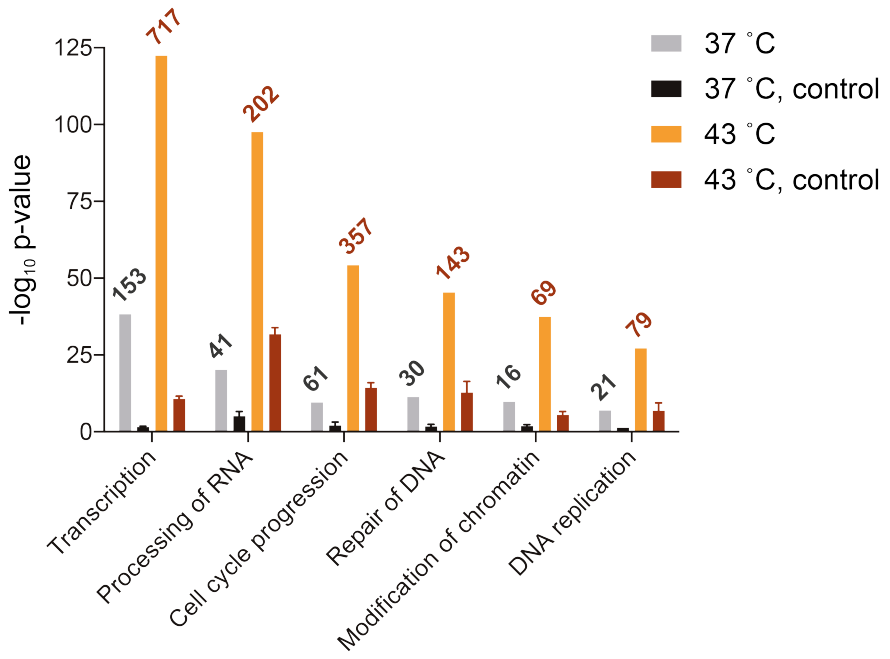

Figure 4.13 | Substrates of 6His-SUMO2^{T90K} are functionally related.

Bar chart showing the significantly overrepresented biological functions of 383 or 2311 sumoylated proteins identified from HEK293$^{6His\text{-}SUMO2\text{-}T90K}$ cells cultured at 37 °C (grey) or 43 °C (orange), respectively. The number of proteins in each functionally related group is depicted above the bars. Triplicate sets of 383 (black) or 2311 (red) randomly selected proteins served as a control. Error bars correspond to standard deviations.

Direct comparisons between the locations of sumoylated or unmodified counterparts of the corresponding proteins revealed a diverse spatial positioning of the non-modified molecules, despite the predominantly nuclear localisation of the SUMO-bound counterparts (Figure 4.12; ordinate). For instance, the non-modified versions of more than a quarter of SUMO substrates were segregated predominantly into the cytoplasmic fraction (Figure 4.12; bottom right section). Interestingly, ~42 % of these proteins were components of ribosomal subunits. The ribosome is a large, cytoplasmically located, macromolecular complex that carries out protein synthesis in every living cell (Noller and Nomura, 1996). The biogenesis of eukaryotic ribosomes is

a tightly regulated process, in which the maturation and assembly of individual components is accompanied by the transport of these pre-ribosomal particles from nucleoli to cytoplasm (Kressler et al., 2010). Interestingly, a connection between the early assembly of pre-ribosomes and sumoylation has been recognised in yeast and human cells, thus explaining the distinct partitioning of SUMO-conjugated molecules into the nuclear subcellular fraction (Panse et al., 2006; Haindl et al., 2008; Finkbeiner et al., 2011).

In response to heat shock, the number of lysines conjugated to SUMO2 increased by more than 13-fold. In total, 8262 modification sites were detected from 2311 proteins extracted from heat-stressed HEK293$^{6His\text{-}SUMO2\text{-}T90K}$ cells (Supplementary table S3; Appendix 6.1). Surprisingly, ~93.7 % of these Gly-Gly-conjugated lysines were detected exclusively after heat shock. Similarly to the distribution of the sumoylated proteome in unperturbed cells, nearly all SUMO2 remnant-containing branched peptides were identified from the nuclear fraction (Figure 4.14; abscissa). However, the distributions of the non-modified counterpart proteins were also shifted towards the nuclear compartment in response to cellular stress (Figure 4.14; ordinate). In particular, the spatial partitioning of the non-modified pool of the heat-induced sumoylated proteome was bimodal, with more than 65 % of the proteins locating into the nuclear fraction and nearly a quarter of them being balanced between the two subcellular compartments (Figure 4.14; ordinate). Furthermore, a subtle overrepresentation of SUMO2 substrates was observed among the 245 proteins, whose location was affected the most in response to heat shock (Figure 4.8 and Figure 4.9). In particular, ~48.6 % of them were sumoylated (Benjamini-Hochberg FDR of Fischer's exact test = 4.2×10^{-9}).

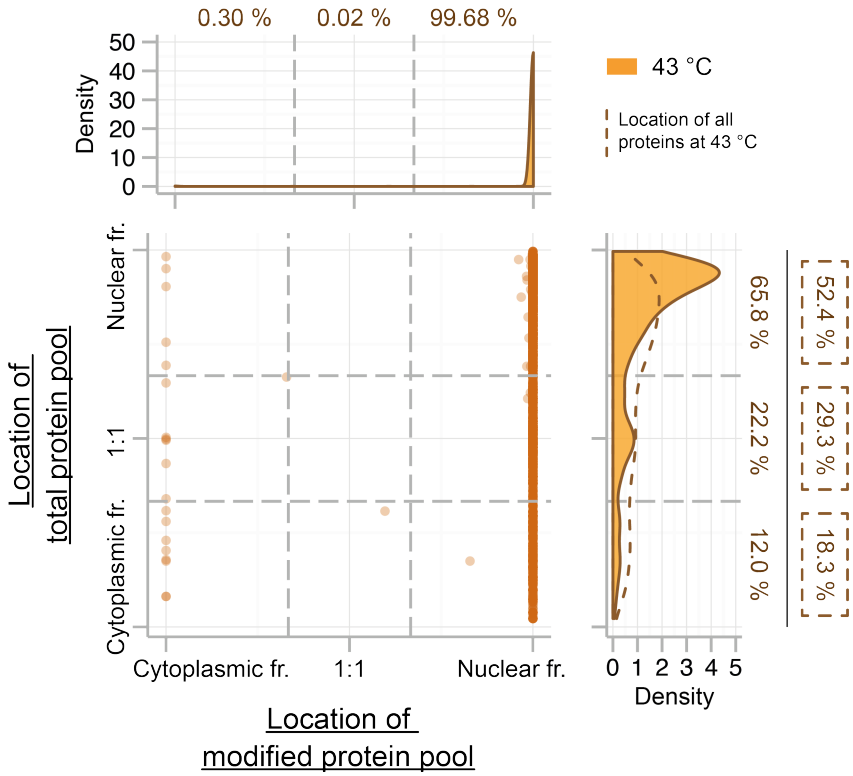

Figure 4.14 | SUMO2-modified proteins segregate into the nuclear fraction of heat-stressed human cells.

Scatter plot with subcellular localisations of individual sumoylated proteins (x-axis) compared to the spatial distributions of the total pools of corresponding proteins (y-axis) in heat shocked HEK293$^{6His\text{-}SUMO2\text{-}T90K}$ cells. The kernel density plots display the probability density function of the subcellular distribution of either SUMO-conjugated (top; filled area) or total (right; filled area) pool of the sumoylated proteome. The probability density function of the localisation of all 5099 proteins quantified in cells cultured at 43 °C is illustrated by the dashed line (right).

Characterisation of the 2311 proteins conjugated to SUMO2 in response to heat shock revealed an enrichment of protein groups associated with the same biological functions as identified for unstressed human cells (Figure 4.13). However, the number of proteins in each functionally related group was increased on average by fourfold. These findings suggest that protein sumoylation functions in specific fundamental cellular processes and the importance of SUMO conjugation is highlighted in cells exposed acute stress-stimuli.

4.2.8. Extent of SUMO2 modification of individual proteins in human cells

Gel electrophoretic pattern of many purified sumoylated conjugates has led to the suggestion that proteins can contain multiple SUMO modification sites (Psakhye and Jentsch, 2012). Indeed, nearly a third (31.8 %) of the sumoylated proteins extracted from unperturbed HEK293$^{6His\text{-}SUMO2\text{-}T90K}$ cells contained multiple lysines, which can be conjugated to SUMO2 (Figure 4.15A; grey). For instance, eleven lysines of ZNF451 were modified in cells cultured at 37 °C (Figure 4.15B). Recently, structural and biochemical studies have revealed that ZNF451 is a SUMO E3 ligase, and can undergo extensive automodification (Cappadocia et al., 2015b; Eisenhardt et al., 2015). Likewise, seven Gly-Gly-lysines were identified from another putative SUMO E3 ligase: Transcription intermediary factor 1-β (TRIM28), which can also undergo autosumoylation (Figure 4.15B) (Ivanov et al., 2007). Many other components of the SUMO conjugation machinery, such as the SUMO-activating enzyme subunit 2 (SAE2) and all SUMO paralogs, contained at least four modification sites at 37 °C (Matunis et al., 1996; Johnson et al., 1997; Lapenta et al., 1997; Desterro et al., 1999). The list of proteins containing multiple SUMO2 modification sites was also enriched in well-established SUMO target proteins, such as Non-POU domain-containing octamer-binding protein (NONO) with ten, Scaffold attachment factor B2 (SAFB2) with seven, and General transcription factor II-I (GTF2I) with five Gly-Gly-lysines (Figure 4.15B) (Becker et al., 2013). Furthermore, four heterogenous nuclear ribonucleoproteins (hnRNPs): M (HNRNPM), C1/C2, (HNRNPC), A0 (HNRNPA0) and A2/B1 (HNRNPA2B1) contained at least four sites, with hnRNP M and hnRNP C1/C2 being well-known substrates of SUMO modification (Vassileva and Matunis, 2004).

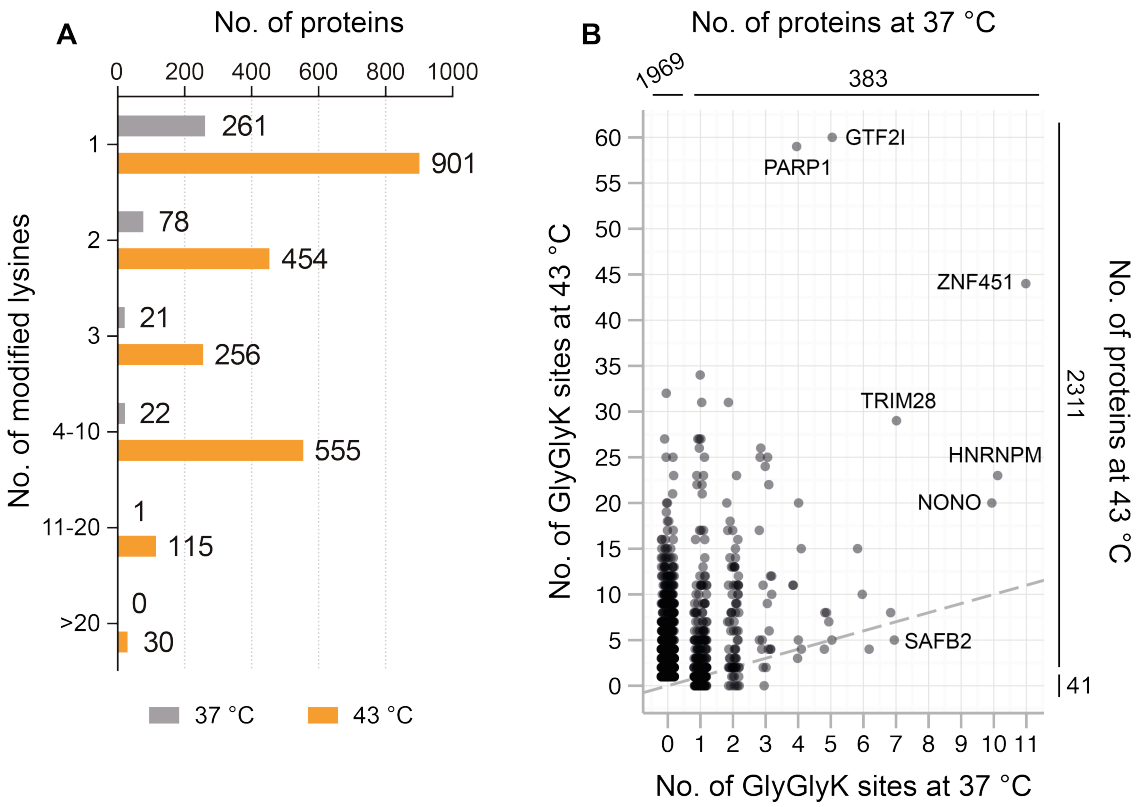

Figure 4.15 | Number of SUMO2 modification sites identified per target protein increases after acute heat stress.

(**A**) Bar chart displaying the number 6His-SUMO2^{T90K} acceptor lysines identified per substrate protein from HEK293$^{6His\text{-}SUMO2\text{-}T90K}$ cells cultured either at 37 °C (grey) or 43 °C (orange). (**B**) Scatter plot illustrating the number of Gly-Gly-lysines modified per target protein in unperturbed (x-axis) or heat-stressed (y-axis) human cells. Proteins with no heat-induced effect on the amount of SUMO2 modification sites map onto grey dashed line. Total number of sumoylated or unmodified proteins in each condition are shown above (37 °C) or next to (43 °C) the scatter plot.

In general, a greater number of SUMO2 modification sites per protein was observed in heat-stressed HEK293$^{6His\text{-}SUMO2\text{-}T90K}$ cells (Figure 4.15B) and the proportion of proteins accommodating multiple Gly-Gly-lysines increased to ~61 % (Figure 4.15A; orange). For instance, GTF2I contained 60, Poly [ADP-ribose] polymerase 1 (PARP1) had 59, and ZNF451 contained 44 sumoylation sites. PARP1 is also a well-established SUMO substrate, whose modification is increased in response to heat shock (Golebiowski et

al., 2009; Martin et al., 2009; Messner et al., 2009). Moreover, 73 proteins with no detectable Gly-Gly-lysines at 37 °C contained more than ten SUMO2-modified residues after heat shock, and for 1969 proteins the SUMO modification was detected in heat-stressed cells only (Figure 4.15B). The number of SUMO2 modification sites was decreased in 51 proteins corresponding to less than 2.2 % of the sumoylated proteome. For instance, the seven SAFB2 lysines modified in unperturbed cells were reduced to five in response to proteotoxic stress. This is consistent with proteomic studies demostrating decreased sumoylation of SAFB2 in human cells exposed to heat or treated with MG132 (Golebiowski et al., 2009; Lamoliatte et al., 2014).

Given the methodological limitations of bottom-up proteomics, for most proteins it was not possible to determine whether the multisite sumoylation occurs simultaneously. However, as Lys-C is inhibited by posttranslational modification of target residues, 71 unique peptide sequences from 39 proteins were identified to contain concurrent sumoylation of adjacent lysines (Supplementary table S4; Appendix 6.1). Most of these multiply modified peptides were detected exclusively after heat shock, and from proteins that contained many Gly-Gly-lysines, such as hnRNPs (A0, A2/B1, C1/C2, F, L and M), PARP1, E3 SUMO-protein ligase NSE2, and all SUMO paralogs. However, multisite modification of SUMO molecules was also detected from cells cultured at 37 °C (Supplementary table S4; Appendix 6.1). Lys^{11} is the major site involved in chain formation on SUMO2 and SUMO3 (Tatham et al., 2001), and the 6His-$SUMO2^{T90K}$-modified Lys^{11} of both paralogs was found in all four analytes. In addition, several combinations of multiply modified peptides of SUMO2 and SUMO3 were detected, including two doubly modified peptides containing Lys^{11} together with either Lys^{7} or Lys^{21}, or a triply modified peptide encompassing Lys^{5}, Lys^{7} and Lys^{11},

suggesting that multiple branching patterns may occur (Supplementary table S4; Appendix 6.1).

4.2.9. Sequence context of SUMO2 modification sites

Initial studies inspecting the amino acid residues targeted for SUMO conjugation suggested that the ability of a lysine to be sumoylated depends on its residence in a conserved sequence environment ψKxE, where ψ and x represent hydrophobic or any residues, respectively (Rodriguez et al., 2001). Consequently, subsequent investigations focused predominantly on these motifs and the determination of SUMO target sites was strongly biased towards lysines residing in the described environment. In contrast, the current resource of SUMO2 modification sites was compiled using high-throughput MS-based proteomics methodology and thus, enabled the detection of modified lysines without previously determined sequence context bias (Choudhary and Mann, 2010). Furthermore, the relatively large size of the acquired data sets made them suitable for the inspection and plausible discovery of novel sequence consensuses utilising robust statistical significance tests.

Therefore, sequence motifs were visualised using pLogo, a web-based software application (O'Shea et al., 2013). Analysis of 612 unique diGly-lysines detected from human cells cultured at 37 °C validated the preferential modification of residues residing in the consensus motif ψKxE (Figure 4.16A) (Rodriguez et al., 2001). However, Val and Ile were the only overrepresented hydrophobic residues N-terminal to the target lysine (position -1). Intriguingly, despite the apparent high frequency of peptides conforming to the consensus motif, the sequence environment of merely

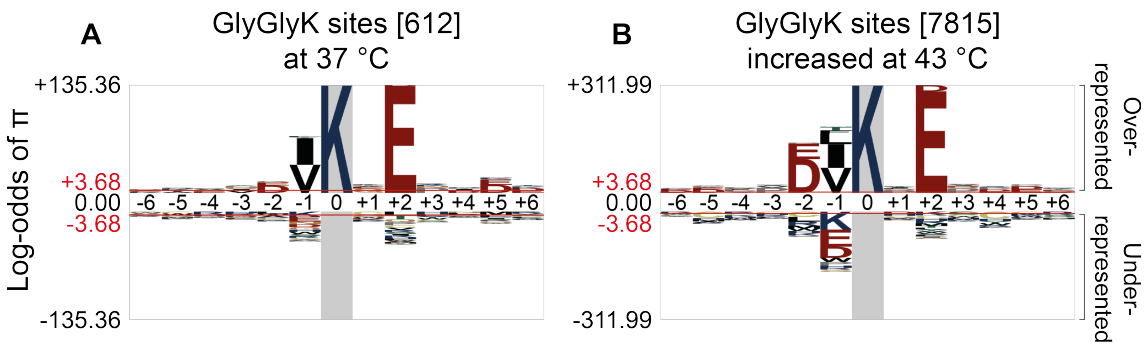

Figure 4.16 | Lysines residing in two distinct sequence motifs are preferentially targeted for SUMO2 conjugation after heat shock.

pLogo representation of the sequence context of: (**A**) 612 lysines modified by 6His-SUMO2^{T90K} in cells cultured at 37 °C, or (**B**) 7815 Gly-Gly-lysines that increased in abundance or were detected exclusively from cells exposed to heat stress. Y-axes correspond to the log-odds of the binomial probability (π) with significance threshold values of 3.68 ($p < 0.05$) marked by red horizontal lines. Note the different scales of y-axes.

26.8 % of the identified lysines matched strictly to this pattern (Figure 4.17A and B). Another 29.6 % of the Gly-Gly-Lys-containing sequences were characterised by a partial consensus motif, which accommodated either a valine or an isoleucine in position -1, or a glutamic acid two amino acids C-terminal to the target lysine (position +2). Surprisingly, nearly a quarter of the sites sumoylated in unperturbed cells did not conform to any significant consensus, thus demonstrating the importance of unbiased identification of sites of SUMO2 modification (Figure 4.17A and B).

Analysis of the 7815 unique Gly-Gly-Lys-containing sequences, which increased in abundance or were detected exclusively after heat shock, uncovered many additional aspects of the widely accepted core motif (Figure 4.16B). Similarly to the sequence environments of lysines modified in unperturbed human cells, ψKxE motif was

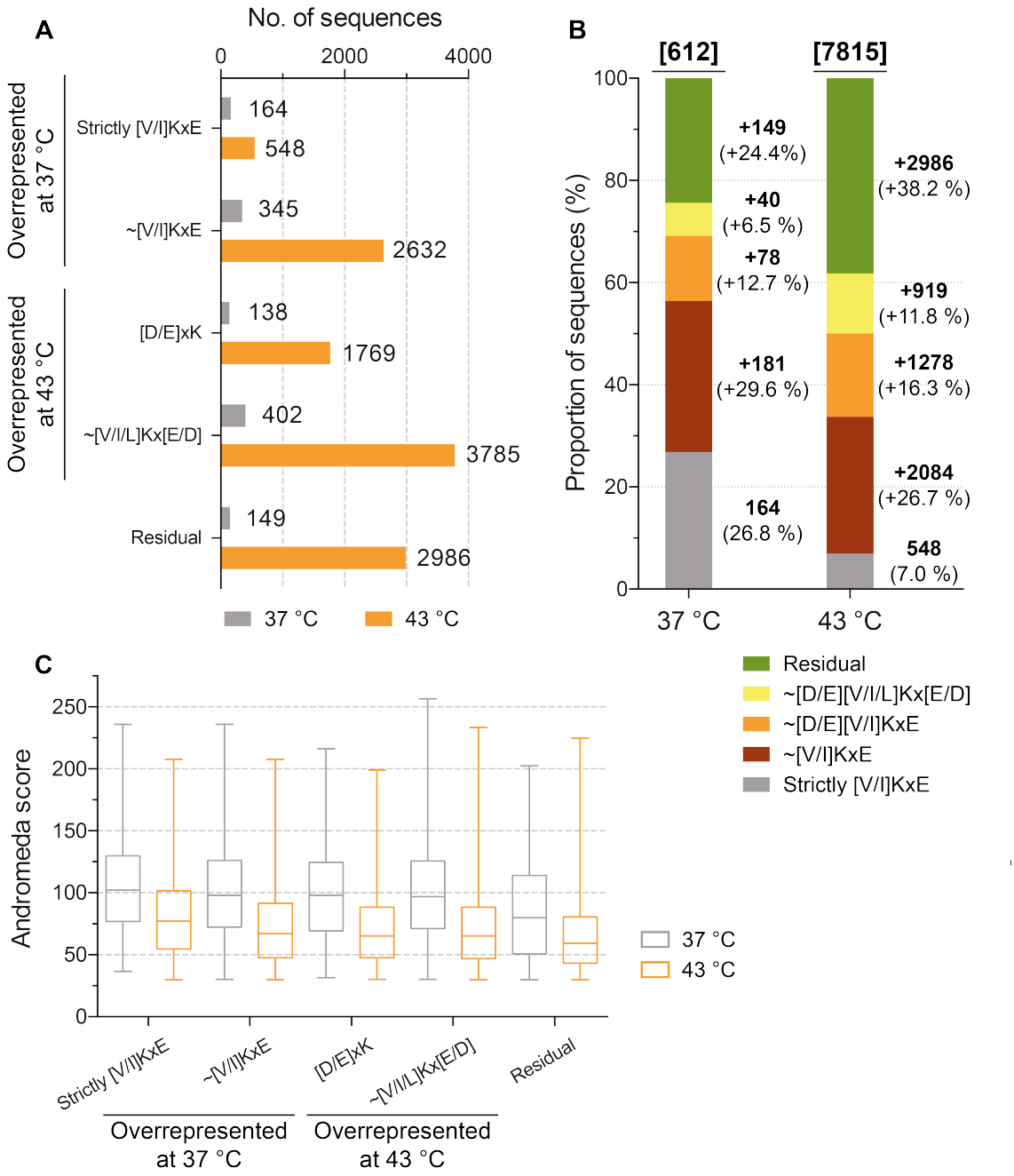

Figure 4.17 | Distribution of the sequence motifs targeted by SUMO2 modification.

››

‹‹ Figure 4.17 | Distribution of the sequence motifs targeted by SUMO2 modification.
(**A**) Bar chart showing the numbers of Gly-Gly-lysines identified from unstressed (grey) or heat-shocked (orange) HEK293$^{6His\text{-}SUMO2\text{-}T90K}$ cells residing in sequence contexts conforming to complete (strictly [V/I]KxE) or partial (~[V/I]KxE) conventional consensus motifs overrepresented in unperturbed cells; or to inverted ([E/D]xK) or partial conventional (~[V/I/L]Kx[E/D]) consensuses preferred exclusively after heat shock. SUMO2 sites that do not conform to any consensus are counted separately (Residual). (**B**) Stacked bar chart demonstrating the cumulative percentage of SUMO2 modification sites conforming to the sequence motifs described in panel A, and identified either from unstressed (37 °C) or heat shocked (43 °C) HEK293$^{6His\text{-}SUMO2\text{-}T90K}$ cells. (**C**) Box plot illustrating the distributions of the Andromeda scores of Gly-Gly-modified peptides grouped by the sequence contexts of the target lysines as in panel A.

overrepresented. However, in addition to Val and Ile, the occurrence of Leu in position -1 was also significantly frequent. Furthermore, the presence of a glutamic acid in position +2 was accompanied by Asp, although, the occurrence of Glu was more than threefold greater (Figure 4.16B).

The preferential modification of lysines in an inverted sumoylation consensus sequence [E/D]xKψ was suggested based on a small-scale proteomic study (Matic et al., 2010). The proposed inverted consensus was overrepresented, but exclusively among the Gly-Gly-Lys-containing sequences modified after the exposure to elevated temperature. Hydrophobic residues did not significantly accompany the target lysines at their C-terminal side. Moreover, the frequency of acidic residues Asp and Glu in position -2 was nearly equal, in contrast to the more prevalent occurrence of Glu in position +2 of the conventional sumoylation consensus motif. Thus, the inverted sequence environment, which is preferentially modified in heat-stressed cells can be characterised simply by an [E/D]xK motif (Figure 4.16B). Interestingly, only 12.7 % of

the target lysines with sequence environments conforming to the inverted consensus motif contained a glutamic acid in position +2. This is consistent with a mixed population of sumoylated sequences matching either to conventional or inverted sumoylation consensus motif.

Although the number or diGly-modified lysines that conformed strictly to the consensus motif ψKxE increased by more than threefold in response to heat stress (Figure 4.17A; strictly [V/I]KxE), these 548 target sites corresponded to only 7.0 % of the overall heat-induced sumoylated lysines (Figure 4.17B). Another 26.7 % and 28.1 % of the modified sites conformed to either partial conventional or inverted sumoylation consensus motif, respectively. The number of SUMO2-modified peptides that did not conform to any significant consensus was increased by ~20-fold, constituting to more than 38 % of all sequence contexts modified in heat-stressed cells. (Figure 4.17A and B). Analysis of the Andromeda scores (the metric for quality of the sequence information) of diGly-modified peptides revealed that the median score of sequences containing heat-induced SUMO2 modification sites was considerably lower (Figure 4.17C). This suggests that their MSMS spectra contained scarce fragmentation information, which could be a consequence of either low abundance or ionisation efficiency. In normal cells, a similar phenomenon was observed only for Gly-Gly-Lys-containing peptides not conforming to any consensus motif (Figure 4.17C).

Dependence of SUMO modification on prior protein phosphorylation has been described for several substrates, such as HSF1 (Hietakangas et al., 2003), Peroxisome proliferator-activated receptor gamma (PPAR-γ) (Yamashita et al., 2004), or MEF2 (Gregoire et al., 2006; Shalizi et al., 2006). These sumoylation sites are likely to fall into

a phosphorylation-dependent sumoylation motif (PDSM), which accommodates a conventional SUMO consensus sequence two residues N-terminal to the proline-directed phosphorylation site (ψKxExxSP) (Hietakangas et al., 2006). In this study, 191 SUMO2 remnant diGly-containing peptides were also phosphorylated (Supplementary table S5; Appendix 6.1). The phosphate-modified residue could be found at nearly any position in proximity to the sumoylation site, but was most prevalent five amino acids C-terminal to the diGly-lysine (Figure 4.18; position +5). Approximately 78.3 % of the sites at position +5 conformed to serine-proline (SP) or threonine-proline (TP) phosphorylation sequence and the amino acid context of 12 Gly-Gly-modified phosphorylated peptides matched strictly to the PDSM motif. A doubly modified peptide of a Transcriptional activator Myb was sumoylated at Lys^{527} and

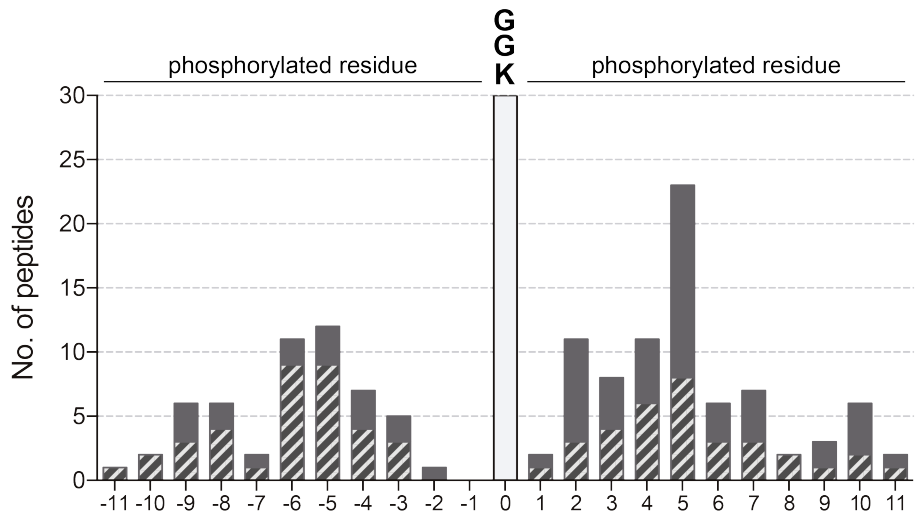

Figure 4.18 | Occurrence of phosphorylated residues in proximity to the sumoylated lysine. Bar chart illustrating the number unique peptide sequences with a phosphorylated residue at the designated position (-11 to -1, or 1 to 11) relative to the SUMO Gly-Gly-remnant modified lysine (position 0). Striped area represents the proportion of sequences with corresponding unphophorylated Gly-Gly-Lys-containing peptides also identified in this study.

phosphorylated at Ser^{532} in a sequence context of IKxExxSP, suggesting phosphorylation-dependent sumoylation (Figure 4.19; top panel). The observation was further confirmed by re-analysing MS data files containing high-resolution MSMS spectra of phosphopeptide-enriched samples using a sequence database of virtual linearised branched peptides (Matic et al., 2008; Matic and Hay, 2012; Mertins et al., 2013). With this approach, the same phosphorylated peptide was also shown to be modified by the tryptic remnant of the wild-type SUMO2 (Figure 4.19; bottom panel). Furthermore, several independent biochemical studies have demonstrated either the sumoylation of Lys^{527} or phosphorylation of Ser^{532} of Myb (Aziz et al., 1995; Bies et al., 2002; Dahle et al., 2003; Sramko et al., 2006; Pani et al., 2008). Although, the dependency of Lys^{527} sumoylation on prior phosphorylation of Ser^{532} has not been explored. Another 12 Gly-Gly-modified peptides were phosphorylated at position +2 or -2 relative to the sumoylated lysine (Figure 4.18). Interestingly, phosphate modification of either of these residues generates a negative charge at a position, which in sequences conforming to conventional or inverted sumoylation consensus motifs are occupied by acidic Glu or Asp residues.

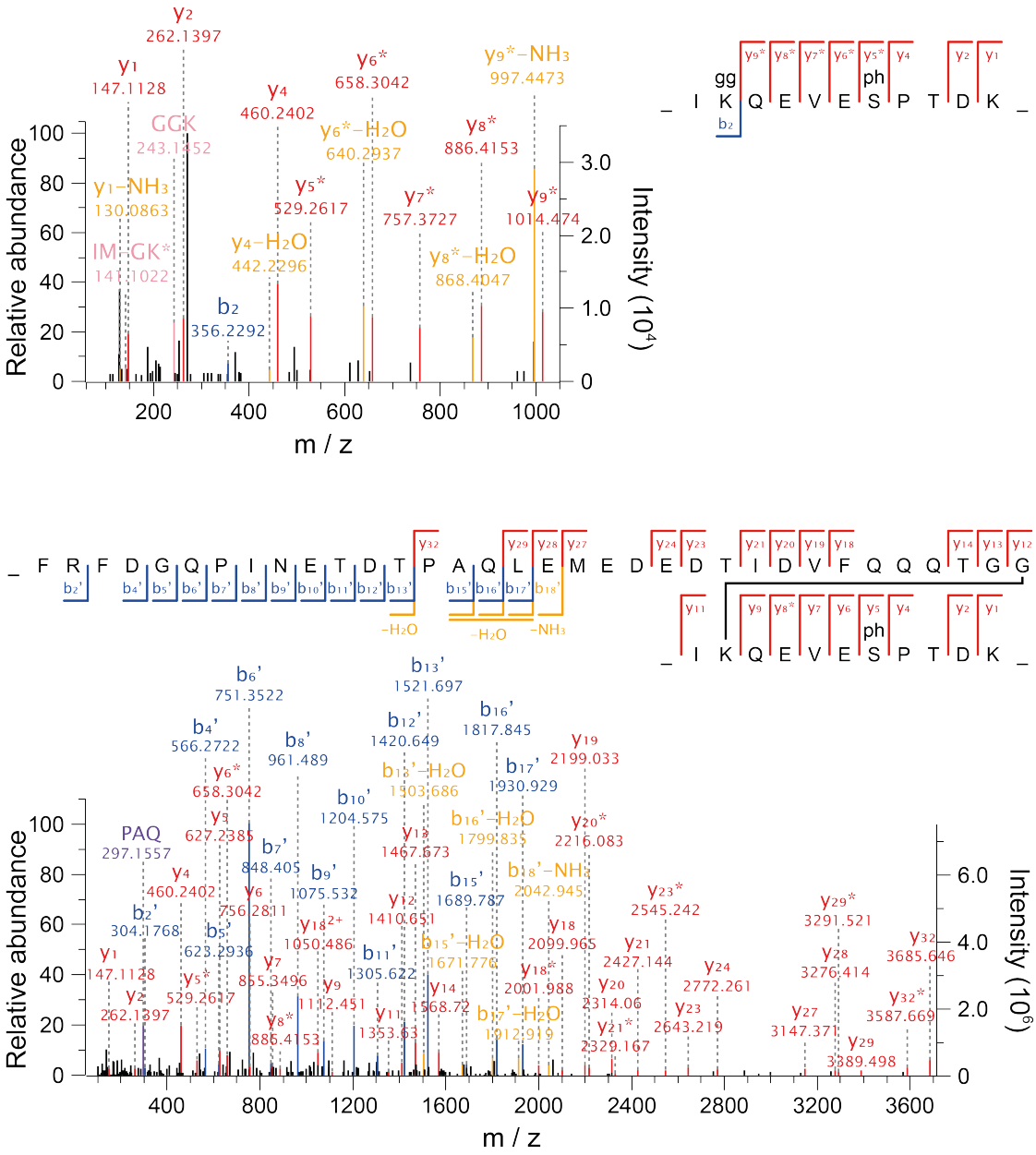

Figure 4.19 | Sequence context of a phosphate-modified sumoylated peptide of Myb conforms to PDSM motif.

Annotated MSMS spectra of a Ser^{532}-phosphorylated peptide of Transcriptional activator Myb with Lys^{527} modified by either Gly-Gly remnant of 6His-SUMO2^{T90K} (top), or C-terminal tryptic fragment of endogenous SUMO2 (bottom). B- (dark blue) and y- (red) ions correspond to fragment ion series extending from the N- or C-terminus of the peptide, respectively. Internal fragments, and ions diagnostic for diGly modification or with a loss of neutral molecule are illustrated in purple, pink or yellow, respectively.

4.2.10. Secondary structure context of the sumoylated lysines

SUMO-conjugating enzyme Ubc9 is able to directly recognise and bind conventional sumoylation consensus motif ψKxE, if presented in an extended conformation (Rodriguez et al., 2001; Sampson et al., 2001; Bernier-Villamor et al., 2002). Consistent with this observation, the majority of Gly-Gly-modified lysines in a sequence environment conforming to partial conventional consensus motif (~[V/I]KxE) were predicted to reside outside the regular secondary structure elements, such as α-helices or β-strands (Figure 4.20; coil). More than 60 % of 6His-SUMO2^{T90K} modification sites resided in coils irrespective of the sequence environment of the target lysine or culturing temperature of the HEK293$^{6His\text{-}SUMO2\text{-}T90K}$ cells. However, the proportion of acceptor lysines found in helices was increased among Gly-Gly-lysines not conforming to any known consensus sequence or in cells exposed to heat stress (Figure 4.20).

Secondary structure elements of a protein can affect the selection between lysines targeted for SUMO conjugation. For instance, Ubc9-dependent *in vitro* sumoylation of Ubiquitin-conjugating enzyme E2 K (E2-25K) occurs exclusively on Lys^{14}, despite the non-consensus sequence context of the target residue (Pichler et al., 2005). This lysine resides in an α-helix and is located between two non-modified consensus sites: Lys^{10} and Lys^{18}. An unstructured peptide corresponding to the N-terminal sequence of E2-25K, however, is preferentially modified at consensus site Lys^{10} (Pichler et al., 2005). Confirming these *in vitro* observations, Lys^{14} of E2-25K alone was conjugated to 6His-SUMO2^{T90K} in the unfractionated sample of HEK293$^{6His\text{-}SUMO2\text{-}T90K}$ cells cultured at 37 °C. However, both Lys^{10} and Lys^{14} were sumoylated in cells cultured at 43 °C (Supplementary table S3), suggesting that the N-terminal α-helix of a proportion of E2-25K could unfold during the 30-minute exposure to elevated temperature.

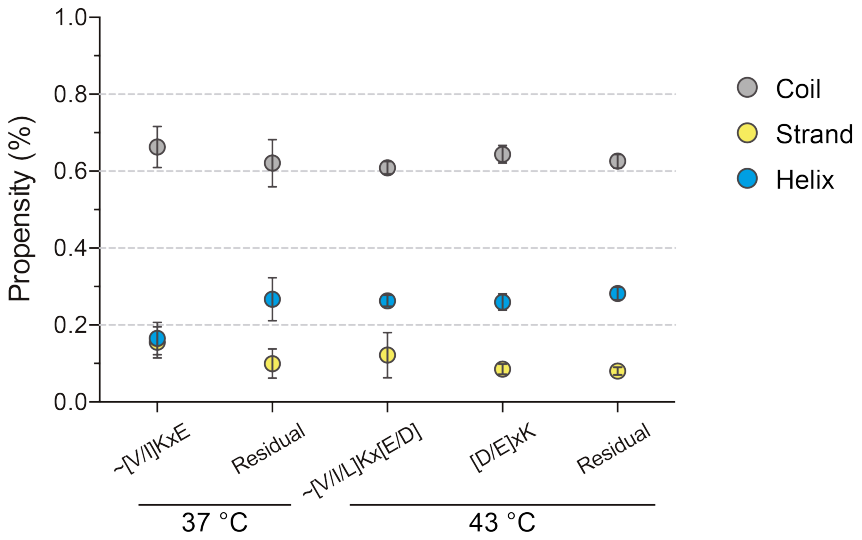

Figure 4.20 | Majority of 6His-SUMO2^{T90K}-conjugated lysines reside outside regular secondary structure elements.

Dot chart demonstrating the propensity of 6His-SUMO2^{T90K}-modified lysines to reside in helical (blue), beta strand (yellow) or coiled (grey) regions of the protein as assigned by JPred4 secondary structure prediction server (Drozdetskiy et al., 2015). Error bars correspond to 95 % confidence intervals.

4.2.11. SUMO2 modification of macromolecular protein complexes

A functional SUMO conjugation pathway is required for the viability of most eukaryotic organisms (Nacerddine et al., 2005). However, lack of phenotypes of particular proteins deficient in SUMO modification has led to the proposal that sumoylation may target macromolecular groups, instead of individual proteins (Psakhye and Jentsch, 2012). As evidenced in Chapter 4.2.7, most SUMO conjugates clustered into functionally related categories and the number of modified proteins in each class was increased considerably in response to proteotoxic stress (Figure 4.13). Gene Ontology cellular component-based annotation of these substrates using PANTHER Classification System revealed that SUMO conjugates were overrepresented among the members of multiprotein complexes (Figure 4.21 and

Figure 4.22) (The Gene Ontology Consortium, 2015; Mi et al., 2016). As expected, several of these protein groups process RNA or function in DNA transcription, such as the spliceosomal small nuclear ribonucleoproteins (snRNPs) or RNA polymerase III transcription factors (Figure 4.21). Moreover, many components of macromolecular complexes involved in DNA replication and repair, or remodelling of chromatin were sumoylated in response to heat stress (Figure 4.22). For instance, 23 of the 30 replisome complex proteins contained Gly-Gly-lysines in cells cultured at 43 °C (Figure 4.22A). Proliferating cell nuclear antigen (PCNA) is one of the best-studied components of the replisome, whose sumoylation at Lys^{164} facilitates the suppression of unscheduled homologous recombination during the replication of *S. cerevisiae* or human DNA (Papouli et al., 2005; Pfander et al., 2005; Gali et al., 2012; Moldovan et al., 2012). Interestingly, Lys^{164}-modified PCNA was one of the few members of the replisome complex conjugated to SUMO2 in unperturbed and heat-stressed $HEK293^{6His\text{-}SUMO2\text{-}T90K}$ cells (Figure 4.22A; Supplementary table S3; Appendix 6.1). Approximately 72 % of the 43 Polycomb group (PcG) epigenetic repressors were sumoylated in cells cultured at 43 °C, with the majority of components containing multiple Gly-Gly-lysines (Figure 4.22B). PcG proteins assemble into two principal multiprotein polycomb repressive complexes (PRC1 and PRC2) that collaborate to silence the expression of numerous target genes (Schwartz and Pirrotta, 2013). Nearly all core components of the two PRC complexes were conjugated to SUMO2 in response to heat stress (Figure 4.22B). Little is known how SUMO conjugation influences the mammalian PcG-dependent repression, except that sumoylated E3 SUMO-protein ligase CBX4 mediates the recruitment of PRC1 to murine DNA (Kang et al., 2010).

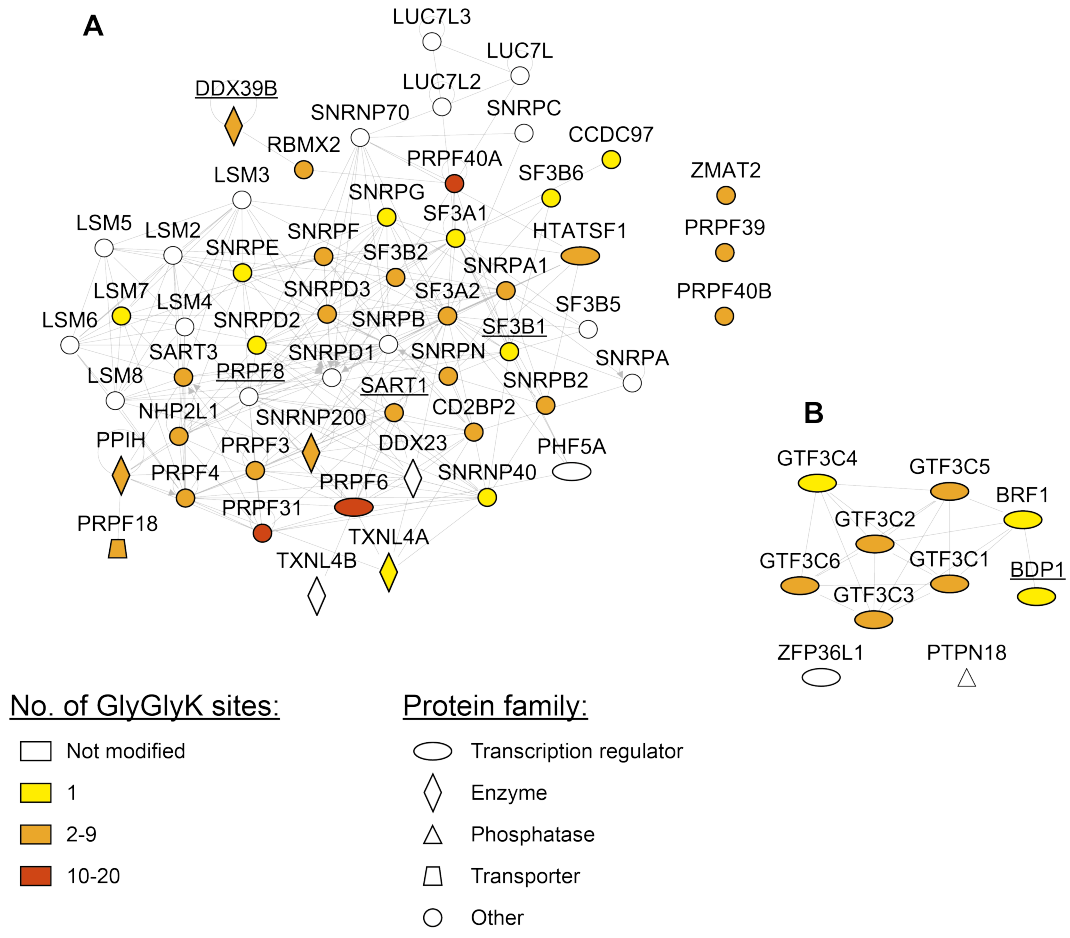

Figure 4.21 | Multiple members of macromolecular protein complexes involved in DNA transcription or RNA processing are conjugated to SUMO2^{T90K} in response to heat stress. Interaction networks of: (**A**) Spliceosomal snRNP complex (accession no. GO:0097525; Bonferroni-corrected p-value 2.37×10^{-12}), or (**B**) RNA polymerase III transcription factor complex (accession no. GO:0090576; Bonferroni-corrected p-value 4.14×10^{-4}). Cellular component-based annotation was retrieved from Gene Ontology database using PANTHER Classification System (The Gene Ontology Consortium, 2015; Mi et al., 2016). As indicated, the colour of the nodes represents the number of sumoylation sites identified from cells cultured at 43 °C (no. of GlyGlyK sites). Proteins modified in unperturbed cells are underlined. The shape of the nodes illustrates the functional class of the protein as provided by Qiagen's Ingenuity Pathway Analysis database (protein family; www.qiagen.com/ingenuity).

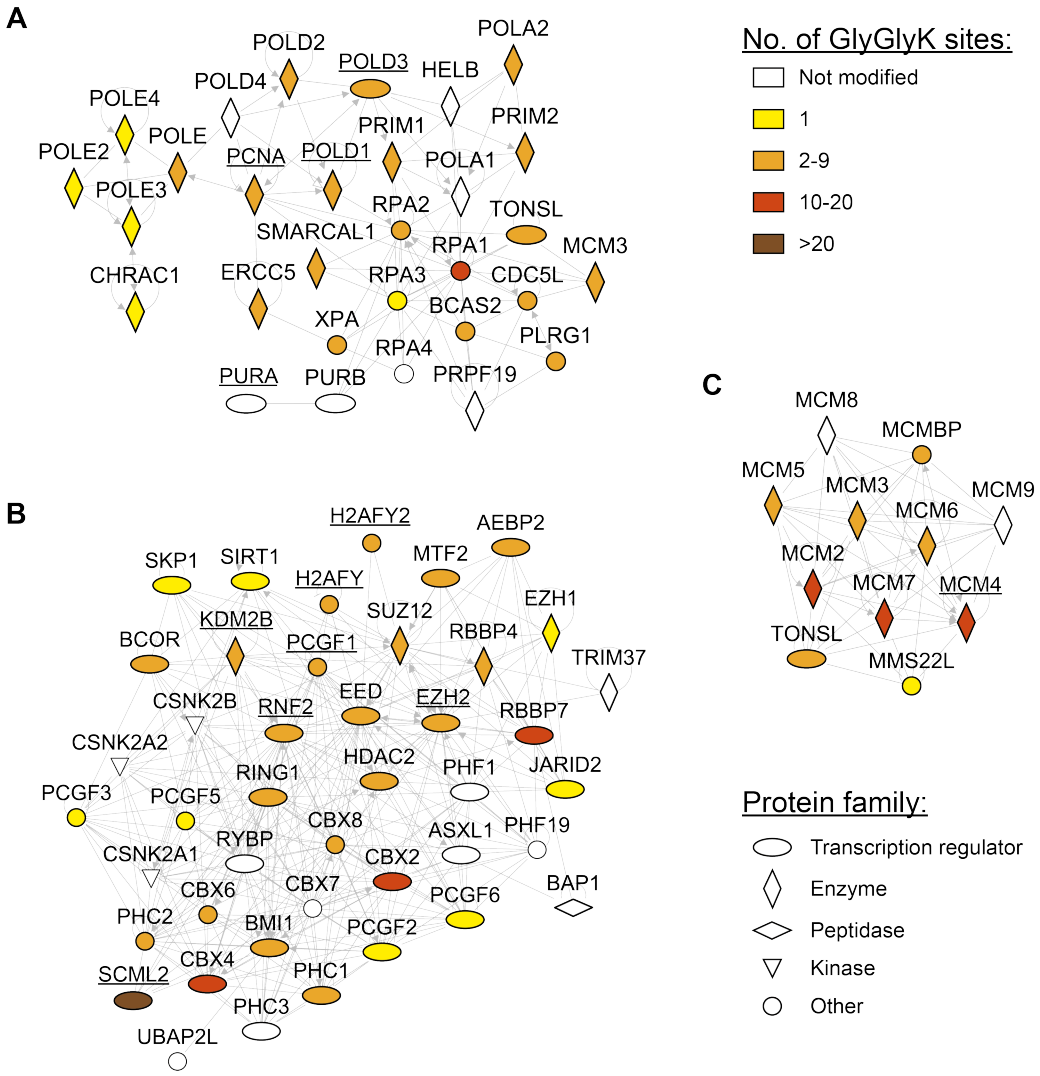

Figure 4.22 | Multiple members of protein complexes involved in DNA replication or chromatin remodelling are conjugated to SUMO2^{T90K} in response to heat stress.

Interaction networks of: (**A**) Replisome (accession no. GO:0030894; Bonferroni-corrected p-value 2.69×10^{-9}), (**B**) PcG protein complex (accession no. GO:0031519; Bonferroni-corrected p-value 2.25×10^{-12}), or (**C**) MCM complex (accession no. GO:0042555; Bonferroni-corrected p-value 7.94×10^{-3}). Cellular component-based annotation was retrieved from Gene Ontology database using PANTHER Classification System (The Gene Ontology Consortium, 2015; Mi et al., 2016). As indicated, the colour of the nodes represents the number of sumoylation sites identified from cells exposed to heat stress (no. of GlyGlyK sites). Proteins modified in unperturbed cells are underlined. The shape of the node illustrates the functional class of the protein as provided by Qiagen's Ingenuity Pathway Analysis database (protein family; www.qiagen.com/ingenuity).

Nine out of 11 proteins of the Minichromosome maintenance (MCM) DNA helicase complex, with all of the six core subunits (Mcm2–7) contained Gly-Gly-lysines in response to proteotoxic stress (Figure 4.22C). In budding yeast, sumoylation of all Mcm2–7 subunits occurs exclusively upon their loading onto chromatin and regulates the initiation of DNA replication by preventing premature origin firing (Wei and Zhao, 2016). The structure of *S. cerevisiae* heterohexameric MCM helicase has been determined (Li et al., 2015), and mapping of the 16 conserved and resolved SUMO2 modification sites onto the surface of the complex showed that none of the SUMO2 molecules would be expected to project into the central channel of the complex. This is consistent with the observation that only chromatin-loaded MCM is sumoylated (Wei and Zhao, 2016). However, several SUMO2-targeted lysines located to the interface of individual subunits (Figure 4.23), suggesting that sumoylation and the association of the subunits are mutually exclusive. The proposal that desumoylation is required prior to the initiation of eukaryotic DNA replication is thus supported by the combined analysis of proteomic and structural data sets (Li et al., 2015; Wei and Zhao, 2016). This example demonstrates how site-specific proteomic studies of sumoylation can support and facilitate the discovery of the function of SUMO2 modification.

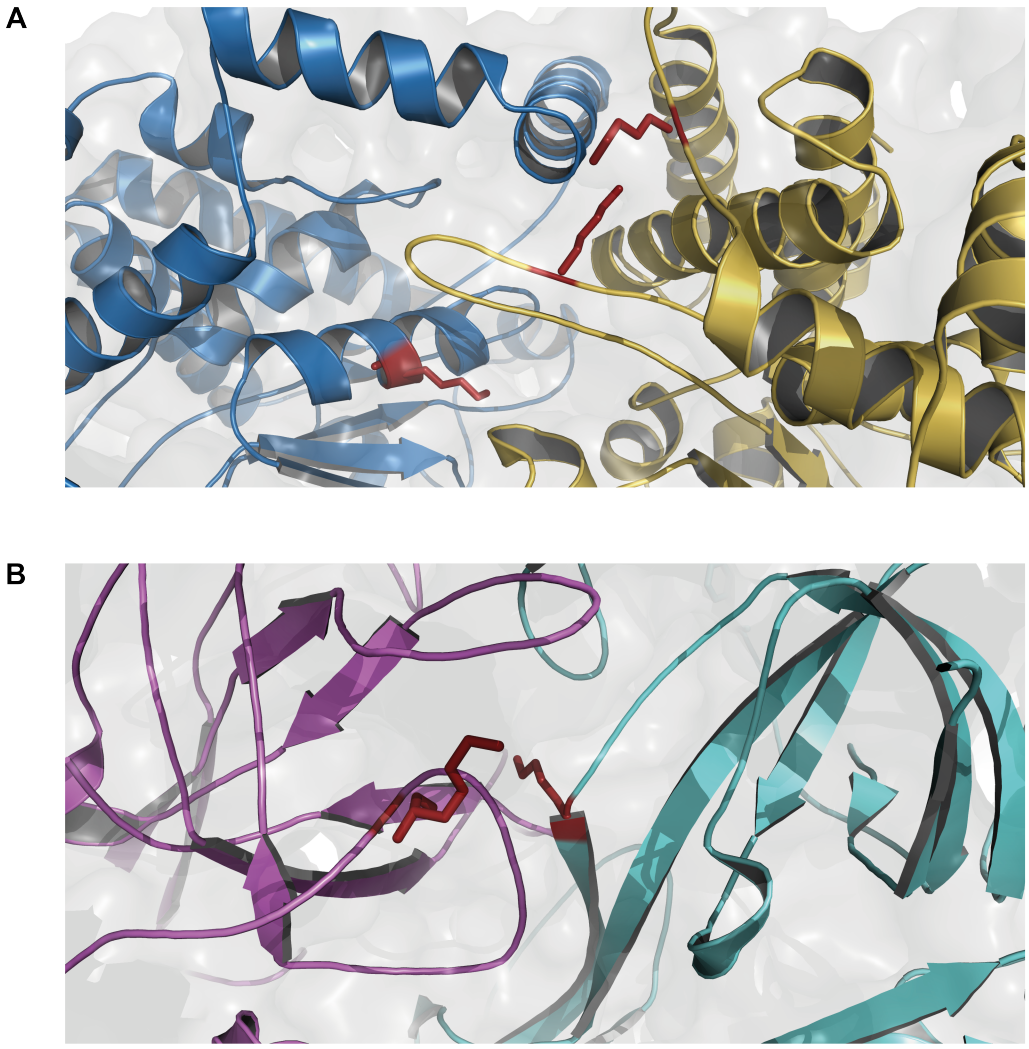

Figure 4.23 | SUMO2^{T90K}-modified lysines locate to the interface of individual subunits of MCM complex.

(**A**) Molecular interface between Mcm5 (blue) and Mcm2 (yellow) with three SUMO2^{T90K} target lysines highlighted in red. (**B**) Molecular interface between Mcm4 (purple) and Mcm7 (cyan) with two SUMO2^{T90K} target lysines highlighted in red. PDB entry: 3JA8 (Li et al., 2015).

4.3. Discussion

4.3.1. Global effects of heat on the HEK293$^{6His\text{-}SUMO2\text{-}T90K}$ proteome

Exposure of cells to elevated temperature triggers a rapid and extensive accumulation of aggregation-prone mis- and unfolded proteins, which serve as stress signals for the activation of a cellular programme known as the heat shock response (Ananthan et al., 1986; Wallace et al., 2015). Stress-induced expression of a specific group of genes encoding molecular chaperones is required to maintain the conformational integrity and the homeostasis of the proteome (proteostasis) (Lindquist, 1986; Ellis, 1987; Hartl, 1996). However, the global production of proteins is repressed at multiple levels of gene regulation via the inhibition of posttranscriptional splicing of RNA, reduced initiation of translation, or halt in translation elongation (Bond, 1988; Spriggs et al., 2010; Shalgi et al., 2013; Shalgi et al., 2014). Meanwhile, the degradation of most eukaryotic proteins is unaffected by hyperthermia (Medicherla and Goldberg, 2008). Consistent with these observations, the absolute abundance of nearly all proteins remained unaltered in HEK293$^{6His\text{-}SUMO2\text{-}T90K}$ cells during a 30-minute heat treatment (Figure 4.7).

Exposure of HEK293$^{6His\text{-}SUMO2\text{-}T90K}$ cells to elevated temperature, however, induced a substantial localisation reorganisation of the proteome, and the abundance of more than a thousand proteins increased in the nuclear or decreased in the cytoplasmic fraction by at least twofold (Figure 4.8 and Figure 4.9). Interestingly, these fold changes were in general produced by a minor proportion of the corresponding protein pool, suggesting that many of those proteins were nearly absent from the nuclear fraction

prior to the heat treatment. Using SILAC-based quantitative proteomics, nuclear accumulation of chaperones, co-chaperones and subunits of the 26S proteasome have been described in HeLa cells exposed to hyperthermia for 2 hours (Raychaudhuri et al., 2014). In the current study, components of the protein ubiquitination pathway were also significantly overrepresented among the 1129 proteins with more than a twofold increase in the nuclear fraction (p-value 9.65×10^{-27}). In particular, the nuclear abundances of 17 (co)-chaperones and 30 subunits of the 26S proteasome were increased in response to hyperthermia. However, the spatial distributions of the members of the two protein families were remarkably different in heat-stressed cells, as chaperones and co-chaperones were mostly nuclear, but the average nuclear proportion of 26S proteasome subunits was only 22 % (Supplementary table S2; Appendix 6.1).

Currently, there are no procedures that enable to obtain subcellular fractions with proteins derived exclusively from nucleus or cytoplasm. A relatively defined fraction of cytoplasmic proteins can be produced through gentle disruption of cells under hypotonic conditions. However, the nuclear fraction will contain substantial amounts of cytoplasmic proteins. Alternatively, cells can be disrupted in the presence of detergents, and the stringent isolation of nuclei will remove most of the contaminating cytoplasmic material. However, soluble nuclear proteins can leach out into the cytoplasmic fraction. In this study, fractionation of HEK293$^{6His\text{-}SUMO2\text{-}T90K}$ cells was optimised to produce a clean preparation of nuclear proteins, including SUMO-conjugated substrates. As a consequence, the cytoplasmic fraction was less defined and contained molecules with predicted nuclear localisation (Figure 4.5 and Figure 4.6). Among these were HSF1 and SUMO2/3, whereas the subsequent immunofluorescence

microscopy-based experiments revealed their predominantly nuclear localisation irrespective of the culturing temperature (Figure 4.10 and Figure 4.11). Notably, both proteins accumulated in subnuclear foci in response to hyperthermia, suggesting that the recruitment of HSF1 or SUMO2/3 to these foci could facilitate their nuclear retention during the subcellular fractionation procedure. Therefore, the extensive heat-induced increase of proteins in the nuclear fraction could be explained by two phenomena:

1) Nuclear proteins could be maintained in the compartment through recruitment to components of the nuclear structure such as chromatin, as observed for HSF1 and SUMO2/3.
2) Cytoplasmic proteins could be translocated to the nucleus.

Importantly, these phenomena are not mutually exclusive. Upon heat shock, YTH domain-containing family protein 2 (YTHDF2) relocates to the nucleus of HeLa cells, where it promotes N^6-adenosine-specific methylation of the 5' untranslated region of mRNA and thus assists the cap-independent translation initiation of selected transcripts (Zhou et al., 2015). In agreement with this observation, the nuclear proportion of YTHDF2 was increased from 30 % to nearly 75 % in HEK293$^{6His\text{-}SUMO2\text{-}T90K}$ cells cultured at 43 °C (Supplementary table S2; Appendix 6.1).

4.3.2. Heat-induced reprogramming of protein sumoylation

Similarly to the rapid aggregation of aberrant proteins, increased conjugation to 6His-SUMO2^{T90K} was observed within minutes of heat treatment, occurring predominantly in the nucleus (Figure 4.1). Prior exposure to proteotoxic stress-stimuli

inhibits heat-induced protein sumoylation and correlates with the increased production of various heat shock proteins (HSPs) (Kurepa et al., 2003; Seifert et al., 2015). Notably, SUMO conjugation is restored by pharmacological inhibition of a Heat shock protein HSP90, suggesting that SUMO modification is repressed by molecular chaperones (Seifert et al., 2015). According to *in vitro* experiments, Lys^{10} in the N-terminal α-helix of Ubiquitin-conjugating enzyme E2 K (E2-25K) is sumoylated when residing in an unstructured peptide (Pichler et al., 2005). In this study, Lys^{10} was conjugated to 6His-$SUMO2^{T90K}$ only in cells cultured at 43 °C, suggesting heat-induced denaturation of E2-25K (Supplementary table S3; Appendix 6.1). Interestingly, sumoylation can promote the solubility of aggregation-prone proteins, such as α-synuclein, Ataxin-7 or Androgen receptor, and suggests that conjugation to SUMO2 may attenuate aggregation of un- and misfolded proteins accumulating after exposure to proteotoxic stress-stimuli, such as heat (Mukherjee et al., 2009; Janer et al., 2010; Krumova et al., 2011).

More than 380 proteins were conjugated to SUMO2 in unperturbed $HEK293^{6His\text{-}SUMO2\text{-}T90K}$ cells. Most of these substrates function in transcription, processing of RNA, remodelling of chromatin, or DNA replication, recombination and repair (Figure 4.13). The rapid increase in SUMO conjugation observed in heat-stressed cells (Figure 4.3) resulted in a substantial group of 2311 sumoylated proteins. Many of these substrates contained multiple SUMO2-modified lysines with the number of protein-specific target sites reaching as many as 60 (Figure 4.15). Moreover, the concurrent modification of adjacent lysines was confirmed for several conjugates (Supplementary table S4; Appendix 6.1). Interestingly, the extensive heat-induced SUMO2 modification appeared to occur on a specific set of substrates with similar function (Figure 4.13) or

affiliated with the same macromolecular complexes (Figure 4.21 and Figure 4.22). Synchronous sumoylation of entire groups of either physically or functionally-related proteins has been suggested to contribute to the formation of macromolecular complexes through multiple SUMO-SIM interactions between individual binding partners (Psakhye and Jentsch, 2012). The substrate selectivity could thus be provided by the recruitment of a limited number of E3 SUMO-protein ligases to these multiprotein complexes (Jentsch and Psakhye, 2013). Supporting this suggestion, an E3 SUMO-protein ligase CBX4 is a member of the Polycomb group of epigenetic repressors that were extensively modified in HEK293$^{6His\text{-}SUMO2\text{-}T90K}$ cells cultured at 43 °C (Figure 4.22) (Kagey et al., 2003). Likewise, SUMO2 modification of chromatin-bound proteins depends on the expression of four genes encoding E3 SUMO-protein ligase PIAS proteins, for which the heat-induced recruitment to active DNA regulatory elements shows a similar kinetics to SUMO2 (Seifert et al., 2015).

The sequence environment of ~56.4 % of the 612 lysines sumoylated in unperturbed HEK293$^{6His\text{-}SUMO2\text{-}T90K}$ cells conformed partially to a SUMO consensus motif [V/I]KxE (Figure 4.16 and Figure 4.17), which is directly recognised by SUMO-conjugating enzyme Ubc9 (Rodriguez et al., 2001; Sampson et al., 2001; Bernier-Villamor et al., 2002). Most of these 345 lysine residues were found outside regular protein secondary structure elements, consistent with the concept of Ubc9 binding the conventional consensus motif if the sequence is presented in an unstructured conformation (Figure 4.20). The number of target lysines with either increased sumoylation status or modified exclusively after heat treatment increased to 7815 in HEK293$^{6His\text{-}SUMO2\text{-}T90K}$ cells cultured at 43 °C (Supplementary table S3; Appendix 6.1). The amino acid context of nearly a half of these sites conformed partially to an altered version of conventional

consensus motif (Figure 4.16 and Figure 4.17; ~[V/I/L]Kx[E/D]). Approximately 22.6 % of them matched to an inverted SUMO consensus motif [E/D]xK, which was overrepresented among the sumoylation sites identified from heat-stressed HEK293$^{6His\text{-}SUMO2\text{-}T90K}$ cells alone. However, the largest heat-induced increase was observed among Gly-Gly-lysines not conforming to any known consensuses and constituted to a nearly 40 % of the SUMO2 modification sites identified after the heat treatment (Figure 4.17B). In addition, the proportion of sites residing in predicted helical secondary structure elements was increased among non-consensus sequences and heat-induced SUMO2 modification sites (Figure 4.20). Whether these alterations to the primary and secondary structure contexts of SUMO2 modification sites were promoted by the relaxed sequence specificity of Ubc9, recruitment of particular E3 SUMO-protein ligases, denaturation of substrates, or all three events simultaneously, requires further investigation.

4.3.3. Future perspectives

The challenge for future research will be the interpretation of large databases of protein localisation or SUMO2 modification sites identified from unperturbed or heat-stressed HEK293$^{6His\text{-}SUMO2\text{-}T90K}$ cells (Supplementary table S3; Appendix 6.1). The current study has described changes in protein localisation in response to hyperthermia, however the biology that underpins these changes has not been explored experimentally. Further investigation, potentially using other technology, should be carried out to reveal whether these changes are induced by biological phenomena, such as nuclear retention or translocation of proteins, or are consequences of the methodology, such as

separation of potential protein aggregates with the nuclear fraction or diffusion of nuclear proteins to the cytoplasmic fraction.

As exemplified in this investigation (Chapter 4.2.11 and Figure 4.23), combined usage of MS-based proteomics, structural and biochemical studies can assist to formulate hypotheses that together with the subsequent *in vivo* experiments could uncover the physiological role of sumoylation of many substrate proteins. Alternatively, these analyses could reveal the instances when protein-group modification is preferred over substrate-specific SUMO modification. The extent of heat-induced sumoylation of individual members of the SUMO conjugation machinery suggests the hyperactivity of the pathway in heat-stressed cells. Therefore, distinguishing between the relevant modifications and bystander events should also be considered.

A number of age-related pathologies, such as Huntington's, Parkinson's or Alzheimer's disease, are characterised by a disruption of proteostasis. SUMO is conjugated to several polypeptides that function in these neurodegenerative diseases (Sarge and Park-Sarge, 2011). The comprehensive database of proteotoxic stress-induced SUMO2 modification sites, provided by this study, could thus be utilised as an *in vivo* model system-based resource to uncover the physiological role of the sumoylation of these substrates in neurodegenerative diseases.

5.

Bibliography

Akerfelt, M., Morimoto, R.I. and Sistonen, L. Heat shock factors: integrators of cell stress, development and lifespan. *Nat Rev Mol Cell Biol* **11**, 545-555 (2010).

Ananthan, J., Goldberg, A.L. and Voellmy, R. Abnormal proteins serve as eukaryotic stress signals and trigger the activation of heat shock genes. *Science* **232**, 522-524 (1986).

Andersen, J.S., Matic, I. and Vertegaal, A.C. Identification of SUMO target proteins by quantitative proteomics. *Methods Mol Biol* **497**, 19-31 (2009).

Anderson, D.D., Eom, J.Y. and Stover, P.J. Competition between sumoylation and ubiquitination of serine hydroxymethyltransferase 1 determines its nuclear localization and its accumulation in the nucleus. *J Biol Chem* **287**, 4790-4799 (2012).

Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T., Harris, M.A., Hill, D.P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J.C., Richardson, J.E., Ringwald, M., Rubin, G.M. and Sherlock, G. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* **25**, 25-29 (2000).

Aziz, N., Miglarese, M.R., Hendrickson, R.C., Shabanowitz, J., Sturgill, T.W., Hunt, D.F. and Bender, T.P. Modulation of c-Myb-induced transcription activation by a phosphorylation site near the negative regulatory domain. *Proc Natl Acad Sci U S A* **92**, 6429-6433 (1995).

Baek, G.H., Cheng, H., Choe, V., Bao, X., Shao, J., Luo, S. and Rao, H. Cdc48: a swiss army knife of cell biology. *J Amino Acids* **2013**, 183421 (2013).

Baler, R., Dahl, G. and Voellmy, R. Activation of human heat shock genes is accompanied by oligomerization, modification, and rapid translocation of heat shock transcription factor HSF1. *Mol Cell Biol* **13**, 2486-2496 (1993).

Bantscheff, M., Schirle, M., Sweetman, G., Rick, J. and Kuster, B. Quantitative mass spectrometry in proteomics: a critical review. *Anal Bioanal Chem* **389**, 1017-1031 (2007).

Bayer, P., Arndt, A., Metzger, S., Mahajan, R., Melchior, F., Jaenicke, R. and Becker, J. Structure determination of the small ubiquitin-related modifier SUMO-1. *J Mol Biol* **280**, 275-286 (1998).

Becker, J., Barysch, S.V., Karaca, S., Dittner, C., Hsiao, H.H., Berriel Diaz, M., Herzig, S., Urlaub, H. and Melchior, F. Detecting endogenous SUMO targets in mammalian cells and tissues. *Nat Struct Mol Biol* **20**, 525-531 (2013).

Bernier-Villamor, V., Sampson, D.A., Matunis, M.J. and Lima, C.D. Structural basis for E2-mediated SUMO conjugation revealed by a complex between ubiquitin-conjugating enzyme Ubc9 and RanGAP1. *Cell* **108**, 345-356 (2002).

Bies, J., Markus, J. and Wolff, L. Covalent attachment of the SUMO-1 protein to the negative regulatory domain of the c-Myb transcription factor modifies its stability and transactivation capacity. *J Biol Chem* **277**, 8999-9009 (2002).

Blomster, H.A., Hietakangas, V., Wu, J., Kouvonen, P., Hautaniemi, S. and Sistonen, L. Novel proteomics strategy brings insight into the prevalence of SUMO-2 target sites. *Mol Cell Proteomics* (2009).

Blomster, H.A., Imanishi, S.Y., Siimes, J., Kastu, J., Morrice, N.A., Eriksson, J.E. and Sistonen, L. In vivo identification of sumoylation sites by a signature tag and cysteine-targeted affinity purification. *J Biol Chem* **285**, 19324-19329 (2010).

Boisvert, F.M., Ahmad, Y., Gierlinski, M., Charriere, F., Lamont, D., Scott, M., Barton, G. and Lamond, A.I. A quantitative spatial proteomics analysis of proteome turnover in human cells. *Mol Cell Proteomics* **11**, M111 011429 (2012).

Bond, U. Heat shock but not other stress inducers leads to the disruption of a sub-set of snRNPs and inhibition of in vitro splicing in HeLa cells. *Embo J* **7**, 3509-3518 (1988).

Bornhorst, J.A. and Falke, J.J. Purification of proteins using polyhistidine affinity tags. *Methods Enzymol* **326**, 245-254 (2000).

Bruderer, R., Tatham, M.H., Plechanovova, A., Matic, I., Garg, A.K. and Hay, R.T. Purification and identification of endogenous polySUMO conjugates. *EMBO Rep* **12**, 142-148 (2011).

Bursomanno, S., Beli, P., Khan, A.M., Minocherhomji, S., Wagner, S.A., Bekker-Jensen, S., Mailand, N., Choudhary, C., Hickson, I.D. and Liu, Y. Proteome-wide analysis of SUMO2 targets in response to pathological DNA replication stress in human cells. *DNA Repair (Amst)* **25**, 84-96 (2015).

Cappadocia, L., Mascle, X.H., Bourdeau, V., Tremblay-Belzile, S., Chaker-Margot, M., Lussier-Price, M., Wada, J., Sakaguchi, K., Aubry, M., Ferbeyre, G. and Omichinski, J.G. Structural and functional characterization of the phosphorylation-dependent interaction between PML and SUMO1. *Structure* **23**, 126-138 (2015a).

Cappadocia, L., Pichler, A. and Lima, C.D. Structural basis for catalytic activation by the human ZNF451 SUMO E3 ligase. *Nat Struct Mol Biol* **22**, 968-975 (2015b).

Cheng, J., Kang, X., Zhang, S. and Yeh, E.T. SUMO-specific protease 1 is essential for stabilization of HIF1alpha during hypoxia. *Cell* **131**, 584-595 (2007).

Choudhary, C. and Mann, M. Decoding signalling networks by mass spectrometry-based proteomics. *Nat Rev Mol Cell Biol* **11**, 427-439 (2010).

Cooper, H.J., Tatham, M.H., Jaffray, E., Heath, J.K., Lam, T.T., Marshall, A.G. and Hay, R.T. Fourier transform ion cyclotron resonance mass spectrometry for the analysis of small ubiquitin-like modifier (SUMO) modification: identification of lysines in RanBP2 and SUMO targeted for modification during the E3 autoSUMOylation reaction. *Anal Chem* **77**, 6310-6319 (2005).

Cotto, J., Fox, S. and Morimoto, R. HSF1 granules: a novel stress-induced nuclear compartment of human cells. *J Cell Sci* **110 (Pt 23)**, 2925-2934 (1997).

Cox, J. and Mann, M. Is proteomics the new genomics? *Cell* **130**, 395-398 (2007).

Cox, J. and Mann, M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat Biotechnol* **26**, 1367-1372 (2008).

Cox, J. and Mann, M. Quantitative, high-resolution proteomics for data-driven systems biology. *Annu Rev Biochem* **80**, 273-299 (2011).

Cox, J., Neuhauser, N., Michalski, A., Scheltema, R.A., Olsen, J.V. and Mann, M. Andromeda: a peptide search engine integrated into the MaxQuant environment. *J Proteome Res* **10**, 1794-1805 (2011).

Cubenas-Potts, C., Srikumar, T., Lee, C., Osula, O., Subramonian, D., Zhang, X.D., Cotter, R.J., Raught, B. and Matunis, M.J. Identification of SUMO-2/3-modified proteins associated with mitotic chromosomes. *Proteomics* **15**, 763-772 (2015).

Dahle, O., Andersen, T.O., Nordgard, O., Matre, V., Del Sal, G. and Gabrielsen, O.S. Transactivation properties of c-Myb are critically dependent on two SUMO-1 acceptor sites that are conjugated in a PIASy enhanced manner. *Eur J Biochem* **270**, 1338-1348 (2003).

Danielsen, J.R., Povlsen, L.K., Villumsen, B.H., Streicher, W., Nilsson, J., Wikstrom, M., Bekker-Jensen, S. and Mailand, N. DNA damage-inducible SUMOylation of HERC2 promotes RNF8 binding via a novel SUMO-binding Zinc finger. *J Cell Biol* **197**, 179-187 (2012).

Demarque, M.D., Nacerddine, K., Neyret-Kahn, H., Andrieux, A., Danenberg, E., Jouvion, G., Bomme, P., Hamard, G., Romagnolo, B., Terris, B., Cumano, A., Barker, N., Clevers, H. and Dejean, A. Sumoylation by Ubc9 regulates the stem cell compartment and structure and function of the intestinal epithelium in mice. *Gastroenterology* **140**, 286-296 (2011).

Desterro, J.M., Rodriguez, M.S. and Hay, R.T. SUMO-1 modification of IκBα inhibits NF-κB activation. *Mol Cell* **2**, 233-239 (1998).

Desterro, J.M., Rodriguez, M.S., Kemp, G.D. and Hay, R.T. Identification of the enzyme required for activation of the small ubiquitin-like protein SUMO-1. *J Biol Chem* **274**, 10618-10624 (1999).

Desterro, J.M., Thomson, J. and Hay, R.T. Ubch9 conjugates SUMO but not ubiquitin. *FEBS Lett* **417**, 297-300 (1997).

Domingues, P., Golebiowski, F., Tatham, M.H., Lopes, A.M., Taggart, A., Hay, R.T. and Hale, B.G. Global Reprogramming of Host SUMOylation during Influenza Virus Infection. *Cell Rep* **13**, 1467-1480 (2015).

Dorval, V. and Fraser, P.E. Small ubiquitin-like modifier (SUMO) modification of natively unfolded proteins tau and alpha-synuclein. *J Biol Chem* **281**, 9919-9924 (2006).

Dorval, V., Mazzella, M.J., Mathews, P.M., Hay, R.T. and Fraser, P.E. Modulation of Abeta generation by small ubiquitin-like modifiers does not require conjugation to target proteins. *Biochem J* **404**, 309-316 (2007).

Drapeau, G.R., Boily, Y. and Houmard, J. Purification and properties of an extracellular protease of Staphylococcus aureus. *J Biol Chem* **247**, 6720-6726 (1972).

Drozdetskiy, A., Cole, C., Procter, J. and Barton, G.J. JPred4: a protein secondary structure prediction server. *Nucleic Acids Res* **43**, W389-394 (2015).

Eckermann, K. SUMO and Parkinson's disease. *Neuromolecular medicine* **15**, 737-759 (2013).

Eidhammer, I., Flikka, K., Martens, L. and Mikalsen, S.O. (2007). Protein Identification and Characterisation by MS, p. 97-118. *In* Computational Methods for Mass Spectrometry Proteomics, John Wiley & Sons, Ltd, Chichester, UK.

Eifler, K. and Vertegaal, A.C. SUMOylation-Mediated Regulation of Cell Cycle Progression and Cancer. *Trends Biochem Sci* **40**, 779-793 (2015).

Eisenhardt, N., Chaugule, V.K., Koidl, S., Droescher, M., Dogan, E., Rettich, J., Sutinen, P., Imanishi, S.Y., Hofmann, K., Palvimo, J.J. and Pichler, A. A new vertebrate SUMO enzyme family reveals insights into SUMO-chain assembly. *Nat Struct Mol Biol* **22**, 959-967 (2015).

Ellis, J. Proteins as molecular chaperones. *Nature* **328**, 378-379 (1987).

Evdokimov, E., Sharma, P., Lockett, S.J., Lualdi, M. and Kuehn, M.R. Loss of SUMO1 in mice affects RanGAP1 localization and formation of PML nuclear bodies, but is not lethal as it can be compensated by SUMO2 or SUMO3. *J Cell Sci* **121**, 4106-4113 (2008).

Everett, R.D., Boutell, C. and Hale, B.G. Interplay between viruses and host sumoylation pathways. *Nat Rev Microbiol* **11**, 400-411 (2013).

Finkbeiner, E., Haindl, M. and Muller, S. The SUMO system controls nucleolar partitioning of a novel mammalian ribosome biogenesis complex. *Embo J* **30**, 1067-1078 (2011).

Flotho, A. and Melchior, F. Sumoylation: a regulatory protein modification in health and disease. *Annu Rev Biochem* **82**, 357-385 (2013).

Fuzesi-Levi, M.G., Ben-Nissan, G., Bianchi, E., Zhou, H., Deery, M.J., Lilley, K.S., Levin, Y. and Sharon, M. Dynamic regulation of the COP9 signalosome in response to DNA damage. *Mol Cell Biol* **34**, 1066-1076 (2014).

Gali, H., Juhasz, S., Morocz, M., Hajdu, I., Fatyol, K., Szukacsov, V., Burkovics, P. and Haracska, L. Role of SUMO modification of human PCNA at stalled replication fork. *Nucleic Acids Res* **40**, 6049-6059 (2012).

Galisson, F., Mahrouche, L., Courcelles, M., Bonneil, E., Meloche, S., Chelbi-Alix, M.K. and Thibault, P. A novel proteomics approach to identify SUMOylated proteins and their modification sites in human cells. *Mol Cell Proteomics* **10**, M110 004796 (2011).

Giansanti, P., Tsiatsiani, L., Low, T.Y. and Heck, A.J. Six alternative proteases for mass spectrometry-based proteomics beyond trypsin. *Nat Protoc* **11**, 993-1006 (2016).

Gibbs-Seymour, I., Oka, Y., Rajendra, E., Weinert, B.T., Passmore, L.A., Patel, K.J., Olsen, J.V., Choudhary, C., Bekker-Jensen, S. and Mailand, N. Ubiquitin-SUMO circuitry controls activated fanconi anemia ID complex dosage in response to DNA damage. *Mol Cell* **57**, 150-164 (2015).

Golebiowski, F., Matic, I., Tatham, M.H., Cole, C., Yin, Y., Nakamura, A., Cox, J., Barton, G.J., Mann, M. and Hay, R.T. System-wide changes to SUMO modifications in response to heat shock. *Sci Signal* **2**, ra24 (2009).

Gong, L., Li, B., Millas, S. and Yeh, E.T. Molecular cloning and characterization of human AOS1 and UBA2, components of the sentrin-activating enzyme complex. *FEBS Lett* **448**, 185-189 (1999).

Gregoire, S., Tremblay, A.M., Xiao, L., Yang, Q., Ma, K., Nie, J., Mao, Z., Wu, Z., Giguere, V. and Yang, X.J. Control of MEF2 transcriptional activity by coordinated phosphorylation and sumoylation. *J Biol Chem* **281**, 4423-4433 (2006).

Haindl, M., Harasim, T., Eick, D. and Muller, S. The nucleolar SUMO-specific protease SENP3 reverses SUMO modification of nucleophosmin and is required for rRNA processing. *EMBO Rep* **9**, 273-279 (2008).

Han, K.J., Jiang, L. and Shu, H.B. Regulation of IRF2 transcriptional activity by its sumoylation. *Biochem Biophys Res Commun* **372**, 772-778 (2008).

Hartl, F.U. Molecular chaperones in cellular protein folding. *Nature* **381**, 571-579 (1996).

Hartl, F.U., Bracher, A. and Hayer-Hartl, M. Molecular chaperones in protein folding and proteostasis. *Nature* **475**, 324-332 (2011).

Hay, R.T. SUMO: a history of modification. *Mol Cell* **18**, 1-12 (2005).

Hecker, C.M., Rabiller, M., Haglund, K., Bayer, P. and Dikic, I. Specification of SUMO1- and SUMO2-interacting motifs. *J Biol Chem* **281**, 16117-16127 (2006).

Hendriks, I.A., D'Souza, R.C., Chang, J.G., Mann, M. and Vertegaal, A.C. System-wide identification of wild-type SUMO-2 conjugation sites. *Nat Commun* **6**, 7289 (2015a).

Hendriks, I.A., D'Souza, R.C., Yang, B., Verlaan-de Vries, M., Mann, M. and Vertegaal, A.C. Uncovering global SUMOylation signaling networks in a site-specific manner. *Nat Struct Mol Biol* **21**, 927-936 (2014).

Hendriks, I.A., Treffers, L.W., Verlaan-de Vries, M., Olsen, J.V. and Vertegaal, A.C. SUMO-2 Orchestrates Chromatin Modifiers in Response to DNA Damage. *Cell Rep* (2015b).

Hendriks, I.A. and Vertegaal, A.C. A comprehensive compilation of SUMO proteomics. *Nat Rev Mol Cell Biol* **17**, 581-595 (2016a).

Hendriks, I.A. and Vertegaal, A.C. A high-yield double-purification proteomics strategy for the identification of SUMO sites. *Nat Protoc* **11**, 1630-1649 (2016b).

Hickey, C.M., Wilson, N.R. and Hochstrasser, M. Function and regulation of SUMO proteases. *Nat Rev Mol Cell Biol* **13**, 755-766 (2012).

Hietakangas, V., Ahlskog, J.K., Jakobsson, A.M., Hellesuo, M., Sahlberg, N.M., Holmberg, C.I., Mikhailov, A., Palvimo, J.J., Pirkkala, L. and Sistonen, L. Phosphorylation of serine 303 is a prerequisite for the stress-inducible SUMO modification of heat shock factor 1. *Mol Cell Biol* **23**, 2953-2968 (2003).

Hietakangas, V., Anckar, J., Blomster, H.A., Fujimoto, M., Palvimo, J.J., Nakai, A. and Sistonen, L. PDSM, a motif for phosphorylation-dependent SUMO modification. *Proc Natl Acad Sci U S A* **103**, 45-50 (2006).

Hobbs, S., Jitrapakdee, S. and Wallace, J.C. Development of a bicistronic vector driven by the human polypeptide chain elongation factor 1alpha promoter for creation of

stable mammalian cell lines that express very high levels of recombinant proteins. *Biochem Biophys Res Commun* **252**, 368-372 (1998).

Hochstrasser, M. Origin and function of ubiquitin-like proteins. *Nature* **458**, 422-429 (2009).

Hochuli, E., Dobeli, H. and Schacher, A. New metal chelate adsorbent selective for proteins and peptides containing neighbouring histidine residues. *J Chromatogr* **411**, 177-184 (1987).

Hornbeck, P.V., Kornhauser, J.M., Tkachev, S., Zhang, B., Skrzypek, E., Murray, B., Latham, V. and Sullivan, M. PhosphoSitePlus: a comprehensive resource for investigating the structure and function of experimentally determined post-translational modifications in man and mouse. *Nucleic Acids Res* **40**, D261-270 (2012).

Houmard, J. and Drapeau, G.R. Staphylococcal protease: a proteolytic enzyme specific for glutamoyl bonds. *Proc Natl Acad Sci U S A* **69**, 3506-3509 (1972).

Hsiao, H.H., Meulmeester, E., Frank, B.T., Melchior, F. and Urlaub, H. "ChopNSpice," a mass spectrometric approach that allows identification of endogenous small ubiquitin-like modifier-conjugated peptides. *Mol Cell Proteomics* **8**, 2664-2675 (2009).

Huang, W.C., Ko, T.P., Li, S.S. and Wang, A.H. Crystal structures of the human SUMO-2 protein at 1.6 A and 1.2 A resolution: implication on the functional differences of SUMO proteins. *Eur J Biochem* **271**, 4114-4122 (2004).

Impens, F., Radoshevich, L., Cossart, P. and Ribet, D. Mapping of SUMO sites and analysis of SUMOylation changes induced by external stimuli. *Proc Natl Acad Sci U S A* **111**, 12432-12437 (2014).

Ivanov, A.V., Peng, H., Yurchenko, V., Yap, K.L., Negorev, D.G., Schultz, D.C., Psulkowski, E., Fredericks, W.J., White, D.E., Maul, G.G., Sadofsky, M.J., Zhou, M.M. and Rauscher, F.J., 3rd PHD domain-mediated E3 ligase activity directs intramolecular

sumoylation of an adjacent bromodomain required for gene silencing. *Mol Cell* **28**, 823-837 (2007).

Janer, A., Werner, A., Takahashi-Fujigasaki, J., Daret, A., Fujigasaki, H., Takada, K., Duyckaerts, C., Brice, A., Dejean, A. and Sittler, A. SUMOylation attenuates the aggregation propensity and cellular toxicity of the polyglutamine expanded ataxin-7. *Hum Mol Genet* **19**, 181-195 (2010).

Jentsch, S. and Psakhye, I. Control of nuclear activities by substrate-selective and protein-group SUMOylation. *Annu Rev Genet* **47**, 167-186 (2013).

Johnson, E.S. and Blobel, G. Ubc9p is the conjugating enzyme for the ubiquitin-like protein Smt3p. *J Biol Chem* **272**, 26799-26802 (1997).

Johnson, E.S., Schwienhorst, I., Dohmen, R.J. and Blobel, G. The ubiquitin-like protein Smt3p is activated for conjugation to other proteins by an Aos1p/Uba2p heterodimer. *Embo J* **16**, 5509-5519 (1997).

Kagey, M.H., Melhuish, T.A. and Wotton, D. The Polycomb Protein Pc2 Is a SUMO E3. *Cell* **113**, 127-137 (2003).

Kahyo, T., Nishida, T. and Yasuda, H. Involvement of PIAS1 in the sumoylation of tumor suppressor p53. *Mol Cell* **8**, 713-718 (2001).

Kaiser, S.E., Riley, B.E., Shaler, T.A., Trevino, R.S., Becker, C.H., Schulman, H. and Kopito, R.R. Protein standard absolute quantification (PSAQ) method for the measurement of cellular ubiquitin pools. *Nat Methods* **8**, 691-696 (2011).

Kamitani, T., Nguyen, H.P. and Yeh, E.T. Preferential modification of nuclear proteins by a novel ubiquitin-like molecule. *J Biol Chem* **272**, 14001-14004 (1997).

Kang, X., Qi, Y., Zuo, Y., Wang, Q., Zou, Y., Schwartz, R.J., Cheng, J. and Yeh, E.T. SUMO-specific protease 2 is essential for suppression of polycomb group protein-mediated gene silencing during embryonic development. *Mol Cell* **38**, 191-201 (2010).

Kim, W., Bennett, E.J., Huttlin, E.L., Guo, A., Li, J., Possemato, A., Sowa, M.E., Rad, R., Rush, J., Comb, M.J., Harper, J.W. and Gygi, S.P. Systematic and quantitative assessment of the ubiquitin-modified proteome. *Mol Cell* **44**, 325-340 (2011).

Kohler, J.B., Tammsalu, T., Jorgensen, M.M., Steen, N., Hay, R.T. and Thon, G. Targeting of SUMO substrates to a Cdc48-Ufd1-Npl4 segregase and STUbL pathway in fission yeast. *Nat Commun* **6**, 8827 (2015).

Kressler, D., Hurt, E. and Bassler, J. Driving ribosome assembly. *Biochim Biophys Acta* **1803**, 673-683 (2010).

Krissinel, E. and Henrick, K. Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Crystallogr D Biol Crystallogr* **60**, 2256-2268 (2004).

Krumova, P., Meulmeester, E., Garrido, M., Tirard, M., Hsiao, H.H., Bossis, G., Urlaub, H., Zweckstetter, M., Kugler, S., Melchior, F., Bahr, M. and Weishaupt, J.H. Sumoylation inhibits alpha-synuclein aggregation and toxicity. *J Cell Biol* **194**, 49-60 (2011).

Kurepa, J., Walker, J.M., Smalle, J., Gosink, M.M., Davis, S.J., Durham, T.L., Sung, D.Y. and Vierstra, R.D. The small ubiquitin-like modifier (SUMO) protein modification system in Arabidopsis. Accumulation of SUMO1 and -2 conjugates is increased by stress. *J Biol Chem* **278**, 6862-6872 (2003).

Labbadia, J. and Morimoto, R.I. Huntington's disease: underlying molecular mechanisms and emerging concepts. *Trends Biochem Sci* **38**, 378-385 (2013).

Labbadia, J. and Morimoto, R.I. The biology of proteostasis in aging and disease. *Annu Rev Biochem* **84**, 435-464 (2015).

LaFerla, F.M., Green, K.N. and Oddo, S. Intracellular amyloid-beta in Alzheimer's disease. *Nat Rev Neurosci* **8**, 499-509 (2007).

Lamoliatte, F., Bonneil, E., Durette, C., Caron-Lizotte, O., Wildemann, D., Zerweck, J., Wenshuk, H. and Thibault, P. Targeted identification of SUMOylation sites in human proteins using affinity enrichment and paralog-specific reporter ions. *Mol Cell Proteomics* **12**, 2536-2550 (2013).

Lamoliatte, F., Caron, D., Durette, C., Mahrouche, L., Maroui, M.A., Caron-Lizotte, O., Bonneil, E., Chelbi-Alix, M.K. and Thibault, P. Large-scale analysis of lysine SUMOylation by SUMO remnant immunoaffinity profiling. *Nat Commun* **5**, 5409 (2014).

Lapenta, V., Chiurazzi, P., van der Spek, P., Pizzuti, A., Hanaoka, F. and Brahe, C. SMT3A, a human homologue of the S. cerevisiae SMT3 gene, maps to chromosome 21qter and defines a novel gene family. *Genomics* **40**, 362-366 (1997).

Lee, L., Sakurai, M., Matsuzaki, S., Arancio, O. and Fraser, P. SUMO and Alzheimer's disease. *Neuromolecular medicine* **15**, 720-736 (2013).

Li, N., Zhai, Y., Zhang, Y., Li, W., Yang, M., Lei, J., Tye, B.K. and Gao, N. Structure of the eukaryotic MCM complex at 3.8 A. *Nature* **524**, 186-191 (2015).

Liebelt, F. and Vertegaal, A.C. Ubiquitin-dependent and independent roles of SUMO in proteostasis. *Am J Physiol Cell Physiol* **311**, C284-296 (2016).

Lin, X., Liang, M., Liang, Y.Y., Brunicardi, F.C., Melchior, F. and Feng, X.H. Activation of transforming growth factor-beta signaling by SUMO-1 modification of tumor suppressor Smad4/DPC4. *J Biol Chem* **278**, 18714-18719 (2003).

Linding, R., Jensen, L.J., Diella, F., Bork, P., Gibson, T.J. and Russell, R.B. Protein disorder prediction: implications for structural proteomics. *Structure* **11**, 1453-1459 (2003).

Lindquist, S. The heat-shock response. *Annu Rev Biochem* **55**, 1151-1191 (1986).

Liu, H.W., Banerjee, T., Guan, X., Freitas, M.A. and Parvin, J.D. The chromatin scaffold protein SAFB1 localizes SUMO-1 to the promoters of ribosomal protein genes to facilitate transcription initiation and splicing. *Nucleic Acids Res* **43**, 3605-3613 (2015).

Liu, H.W., Zhang, J., Heine, G.F., Arora, M., Gulcin Ozer, H., Onti-Srinivasan, R., Huang, K. and Parvin, J.D. Chromatin modification by SUMO-1 stimulates the promoters of translation machinery genes. *Nucleic Acids Res* **40**, 10172-10186 (2012).

Luo, H.B., Xia, Y.Y., Shu, X.J., Liu, Z.C., Feng, Y., Liu, X.H., Yu, G., Yin, G., Xiong, Y.S., Zeng, K., Jiang, J., Ye, K., Wang, X.C. and Wang, J.Z. SUMOylation at K340 inhibits tau degradation through deregulating its phosphorylation and ubiquitination. *Proc Natl Acad Sci U S A* **111**, 16586-16591 (2014).

Mahajan, R., Delphin, C., Guan, T., Gerace, L. and Melchior, F. A small ubiquitin-related polypeptide involved in targeting RanGAP1 to nuclear pore complex protein RanBP2. *Cell* **88**, 97-107 (1997).

Manza, L.L., Stamer, S.L., Ham, A.J., Codreanu, S.G. and Liebler, D.C. Sample preparation and digestion for proteomic analyses using spin filters. *Proteomics* **5**, 1742-1745 (2005).

Martin, N., Schwamborn, K., Schreiber, V., Werner, A., Guillier, C., Zhang, X.D., Bischof, O., Seeler, J.S. and Dejean, A. PARP-1 transcriptional activity is regulated by sumoylation upon heat shock. *Embo J* **28**, 3534-3548 (2009).

Matafora, V., D'Amato, A., Mori, S., Blasi, F. and Bachi, A. Proteomics analysis of nucleolar SUMO-1 target proteins upon proteasome inhibition. *Mol Cell Proteomics* **8**, 2243-2255 (2009).

Matic, I. and Hay, R.T. Detection and quantitation of SUMO chains by mass spectrometry. *Methods Mol Biol* **832**, 239-247 (2012).

Matic, I., Jaffray, E.G., Oxenham, S.K., Groves, M.J., Barratt, C.L., Tauro, S., Stanley-Wall, N.R. and Hay, R.T. Absolute SILAC-compatible expression strain allows Sumo-2 copy number determination in clinical samples. *J Proteome Res* **10**, 4869-4875 (2011).

Matic, I., Schimmel, J., Hendriks, I.A., van Santen, M.A., van de Rijke, F., van Dam, H., Gnad, F., Mann, M. and Vertegaal, A.C. Site-specific identification of SUMO-2 targets in cells reveals an inverted SUMOylation motif and a hydrophobic cluster SUMOylation motif. *Mol Cell* **39**, 641-652 (2010).

Matic, I., van Hagen, M., Schimmel, J., Macek, B., Ogg, S.C., Tatham, M.H., Hay, R.T., Lamond, A.I., Mann, M. and Vertegaal, A.C. In vivo identification of human small ubiquitin-like modifier polymerization sites by high accuracy mass spectrometry and an in vitro to in vivo strategy. *Mol Cell Proteomics* **7**, 132-144 (2008).

Matunis, M.J., Coutavas, E. and Blobel, G. A novel ubiquitin-like modification modulates the partitioning of the Ran-GTPase-activating protein RanGAP1 between the cytosol and the nuclear pore complex. *J Cell Biol* **135**, 1457-1470 (1996).

Medicherla, B. and Goldberg, A.L. Heat shock and oxygen radicals stimulate ubiquitin-dependent degradation mainly of newly synthesized proteins. *J Cell Biol* **182**, 663-673 (2008).

Mendler, L., Braun, T. and Muller, S. The Ubiquitin-Like SUMO System and Heart Function: From Development to Disease. *Circ Res* **118**, 132-144 (2016).

Mertins, P., Qiao, J.W., Patel, J., Udeshi, N.D., Clauser, K.R., Mani, D.R., Burgess, M.W., Gillette, M.A., Jaffe, J.D. and Carr, S.A. Integrated proteomic analysis of post-translational modifications by serial enrichment. *Nat Methods* **10**, 634-637 (2013).

Messner, S., Schuermann, D., Altmeyer, M., Kassner, I., Schmidt, D., Schar, P., Muller, S. and Hottiger, M.O. Sumoylation of poly(ADP-ribose) polymerase 1 inhibits its acetylation and restrains transcriptional coactivator function. *FASEB J* **23**, 3978-3989 (2009).

Meyer, H., Bug, M. and Bremer, S. Emerging functions of the VCP/p97 AAA-ATPase in the ubiquitin system. *Nat Cell Biol* **14**, 117-123 (2012).

Meyer, J.G., Kim, S., Maltby, D.A., Ghassemian, M., Bandeira, N. and Komives, E.A. Expanding proteome coverage with orthogonal-specificity alpha-lytic proteases. *Mol Cell Proteomics* **13**, 823-835 (2014).

Mi, H., Poudel, S., Muruganujan, A., Casagrande, J.T. and Thomas, P.D. PANTHER version 10: expanded protein families and functions, and analysis tools. *Nucleic Acids Res* **44**, D336-342 (2016).

Moldovan, G.L., Dejsuphong, D., Petalcorin, M.I., Hofmann, K., Takeda, S., Boulton, S.J. and D'Andrea, A.D. Inhibition of homologous recombination by the PCNA-interacting protein PARI. *Mol Cell* **45**, 75-86 (2012).

Mukherjee, S., Thomas, M., Dadgar, N., Lieberman, A.P. and Iniguez-Lluhi, J.A. Small ubiquitin-like modifier (SUMO) modification of the androgen receptor attenuates polyglutamine-mediated aggregation. *J Biol Chem* **284**, 21296-21306 (2009).

Mukhopadhyay, D. and Dasso, M. The fate of metaphase kinetochores is weighed in the balance of SUMOylation during S phase. *Cell Cycle* **9**, 3194-3201 (2010).

Mullan, M., Crawford, F., Axelman, K., Houlden, H., Lilius, L., Winblad, B. and Lannfelt, L. A pathogenic mutation for probable Alzheimer's disease in the APP gene at the N-terminus of beta-amyloid. *Nat Genet* **1**, 345-347 (1992).

Muller, S., Berger, M., Lehembre, F., Seeler, J.S., Haupt, Y. and Dejean, A. c-Jun and p53 Activity Is Modulated by SUMO-1 Modification. *J Biol Chem* **275**, 13321-13329 (2000).

Muller, S., Hoege, C., Pyrowolakis, G. and Jentsch, S. SUMO, ubiquitin's mysterious cousin. *Nat Rev Mol Cell Biol* **2**, 202-210 (2001).

Nacerddine, K., Lehembre, F., Bhaumik, M., Artus, J., Cohen-Tannoudji, M., Babinet, C., Pandolfi, P.P. and Dejean, A. The SUMO pathway is essential for nuclear integrity and chromosome segregation in mice. *Dev Cell* **9**, 769-779 (2005).

Neyret-Kahn, H., Benhamed, M., Ye, T., Le Gras, S., Cossec, J.C., Lapaquette, P., Bischof, O., Ouspenskaia, M., Dasso, M., Seeler, J., Davidson, I. and Dejean, A. Sumoylation at chromatin governs coordinated repression of a transcriptional program essential for cell growth and proliferation. *Genome Res* **23**, 1563-1579 (2013).

Nie, M. and Boddy, M.N. Cooperativity of the SUMO and Ubiquitin Pathways in Genome Stability. *Biomolecules* **6**, 14 (2016).

Nielsen, M.L., Vermeulen, M., Bonaldi, T., Cox, J., Moroder, L. and Mann, M. Iodoacetamide-induced artifact mimics ubiquitination in mass spectrometry. *Nat Methods* **5**, 459-460 (2008).

Nilsson, T., Mann, M., Aebersold, R., Yates, J.R., 3rd, Bairoch, A. and Bergeron, J.J. Mass spectrometry in high-throughput proteomics: ready for the big time. *Nat Methods* **7**, 681-685 (2010).

Nishida, T., Kaneko, F., Kitagawa, M. and Yasuda, H. Characterization of a novel mammalian SUMO-1/Smt3-specific isopeptidase, a homologue of rat axam, which is an axin-binding protein promoting beta-catenin degradation. *J Biol Chem* **276**, 39060-39066 (2001).

Niskanen, E.A., Malinen, M., Sutinen, P., Toropainen, S., Paakinaho, V., Vihervaara, A., Joutsen, J., Kaikkonen, M.U., Sistonen, L. and Palvimo, J.J. Global SUMOylation on active chromatin is an acute heat stress response restricting transcription. *Genome Biol* **16**, 153 (2015).

Noller, H.F. and Nomura, M. (1996). Ribosomes, p. 167-186. *In* Escherichia coli and Salmonella, ASM Press, Washington, D. C.

O'Rourke, J.G., Gareau, J.R., Ochaba, J., Song, W., Rasko, T., Reverter, D., Lee, J., Monteys, A.M., Pallos, J., Mee, L., Vashishtha, M., Apostol, B.L., Nicholson, T.P., Illes, K., Zhu, Y.Z., Dasso, M., Bates, G.P., Difiglia, M., Davidson, B., Wanker, E.E., Marsh, J.L., Lima, C.D., Steffan, J.S. and Thompson, L.M. SUMO-2 and PIAS1 modulate insoluble mutant huntingtin protein accumulation. *Cell Rep* **4**, 362-375 (2013).

O'Shea, J.P., Chou, M.F., Quader, S.A., Ryan, J.K., Church, G.M. and Schwartz, D. pLogo: a probabilistic approach to visualizing sequence motifs. *Nat Methods* **10**, 1211-1212 (2013).

Okuma, T., Honda, R., Ichikawa, G., Tsumagari, N. and Yasuda, H. In vitro SUMO-1 modification requires two enzymatic steps, E1 and E2. *Biochem Biophys Res Commun* **254**, 693-698 (1999).

Olsen, J.V., Ong, S.E. and Mann, M. Trypsin cleaves exclusively C-terminal to arginine and lysine residues. *Mol Cell Proteomics* **3**, 608-614 (2004).

Ong, S.E., Blagoev, B., Kratchmarova, I., Kristensen, D.B., Steen, H., Pandey, A. and Mann, M. Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Mol Cell Proteomics* **1**, 376-386 (2002).

Orr, H.T. and Zoghbi, H.Y. Trinucleotide repeat disorders. *Annu Rev Neurosci* **30**, 575-621 (2007).

Palmer, A., Rivett, A.J., Thomson, S., Hendil, K.B., Butcher, G.W., Fuertes, G. and Knecht, E. Subpopulations of proteasomes in rat liver nuclei, microsomes and cytosol. *Biochem J* **316 (Pt 2)**, 401-407 (1996).

Pani, E., Menigatti, M., Schubert, S., Hess, D., Gerrits, B., Klempnauer, K.H. and Ferrari, S. Pin1 interacts with c-Myb in a phosphorylation-dependent manner and regulates its transactivation activity. *Biochim Biophys Acta* **1783**, 1121-1128 (2008).

Panse, V.G., Kressler, D., Pauli, A., Petfalski, E., Gnadig, M., Tollervey, D. and Hurt, E. Formation and Nuclear Export of Preribosomes Are Functionally Linked to the Small-Ubiquitin-Related Modifier Pathway. *Traffic* **7**, 1311-1321 (2006).

Papouli, E., Chen, S., Davies, A.A., Huttner, D., Krejci, L., Sung, P. and Ulrich, H.D. Crosstalk between SUMO and ubiquitin on PCNA is mediated by recruitment of the helicase Srs2p. *Mol Cell* **19**, 123-133 (2005).

Pedrioli, P.G., Raught, B., Zhang, X.D., Rogers, R., Aitchison, J., Matunis, M. and Aebersold, R. Automated identification of SUMOylation sites using mass spectrometry and SUMmOn pattern recognition software. *Nat Methods* **3**, 533-539 (2006).

Pelisch, F., Gerez, J., Druker, J., Schor, I.E., Munoz, M.J., Risso, G., Petrillo, E., Westman, B.J., Lamond, A.I., Arzt, E. and Srebrow, A. The serine/arginine-rich protein SF2/ASF regulates protein sumoylation. *Proc Natl Acad Sci U S A* **107**, 16119-16124 (2010).

Pelisch, F., Tammsalu, T., Wang, B., Jaffray, E.G., Gartner, A. and Hay, R.T. A SUMO-Dependent Protein Network Regulates Chromosome Congression during Oocyte Meiosis. *Mol Cell* (2016).

Pfander, B., Moldovan, G.L., Sacher, M., Hoege, C. and Jentsch, S. SUMO-modified PCNA recruits Srs2 to prevent recombination during S phase. *Nature* **436**, 428-433 (2005).

Pichler, A., Gast, A., Seeler, J.S., Dejean, A. and Melchior, F. The nucleoporin RanBP2 has SUMO1 E3 ligase activity. *Cell* **108**, 109-120 (2002).

Pichler, A., Knipscheer, P., Oberhofer, E., van Dijk, W.J., Korner, R., Olsen, J.V., Jentsch, S., Melchior, F. and Sixma, T.K. SUMO modification of the ubiquitin-conjugating enzyme E2-25K. *Nat Struct Mol Biol* **12**, 264-269 (2005).

Pichler, A., Knipscheer, P., Saitoh, H., Sixma, T.K. and Melchior, F. The RanBP2 SUMO E3 ligase is neither HECT- nor RING-type. *Nat Struct Mol Biol* **11**, 984-991 (2004).

Pilla, E., Moller, U., Sauer, G., Mattiroli, F., Melchior, F. and Geiss-Friedlander, R. A novel SUMO1-specific interacting motif in dipeptidyl peptidase 9 (DPP9) that is important for enzymatic regulation. *J Biol Chem* **287**, 44320-44329 (2012).

Potts, P.R. and Yu, H. Human MMS21/NSE2 is a SUMO ligase required for DNA repair. *Mol Cell Biol* **25**, 7021-7032 (2005).

Poulsen, J.W., Madsen, C.T., Young, C., Poulsen, F.M. and Nielsen, M.L. Using guanidine-hydrochloride for fast and efficient protein digestion and single-step affinity-purification mass spectrometry. *J Proteome Res* **12**, 1020-1030 (2013).

Psakhye, I. and Jentsch, S. Protein group modification and synergy in the SUMO pathway as exemplified in DNA repair. *Cell* **151**, 807-820 (2012).

R Core Team (2015). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org/.

Raijmakers, R., Neerincx, P., Mohammed, S. and Heck, A.J. Cleavage specificities of the brother and sister proteases Lys-C and Lys-N. *Chem Commun (Camb)* **46**, 8827-8829 (2010).

Rappsilber, J., Mann, M. and Ishihama, Y. Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. *Nat Protoc* **2**, 1896-1906 (2007).

Raychaudhuri, S., Loew, C., Korner, R., Pinkert, S., Theis, M., Hayer-Hartl, M., Buchholz, F. and Hartl, F.U. Interplay of acetyltransferase EP300 and the proteasome system in regulating heat shock transcription factor 1. *Cell* **156**, 975-985 (2014).

Reverter, D. and Lima, C.D. Insights into E3 ligase activity revealed by a SUMO-RanGAP1-Ubc9-Nup358 complex. *Nature* **435**, 687-692 (2005).

Rivett, A.J., Palmer, A. and Knecht, E. Electron microscopic localization of the multicatalytic proteinase complex in rat liver and in cultured cells. *J Histochem Cytochem* **40**, 1165-1172 (1992).

Rodriguez, M.S., Dargemont, C. and Hay, R.T. SUMO-1 conjugation in vivo requires both a consensus modification motif and nuclear targeting. *J Biol Chem* **276**, 12654-12659 (2001).

Rodriguez, M.S., Desterro, J.M., Lain, S., Midgley, C.A., Lane, D.P. and Hay, R.T. SUMO-1 modification activates the transcriptional response of p53. *Embo J* **18**, 6455-6461 (1999).

RStudio Team (2015). RStudio: Integrated Development for R. RStudio, Inc., Boston, MA, US. http://www.rstudio.com/.

Ruepp, A., Brauner, B., Dunger-Kaltenbach, I., Frishman, G., Montrone, C., Stransky, M., Waegele, B., Schmidt, T., Doudieu, O.N., Stumpflen, V. and Mewes, H.W. CORUM: the comprehensive resource of mammalian protein complexes. *Nucleic Acids Res* **36**, D646-650 (2008).

Sachdev, S., Bruhn, L., Sieber, H., Pichler, A., Melchior, F. and Grosschedl, R. PIASy, a nuclear matrix-associated SUMO E3 ligase, represses LEF1 activity by sequestration into nuclear bodies. *Genes Dev* **15**, 3088-3103 (2001).

Sadygov, R.G., Cociorva, D. and Yates, J.R., 3rd Large-scale database searching using tandem mass spectra: looking up the answer in the back of the book. *Nat Methods* **1**, 195-202 (2004).

Saitoh, H. and Hinchey, J. Functional heterogeneity of small ubiquitin-related protein modifiers SUMO-1 versus SUMO-2/3. *J Biol Chem* **275**, 6252-6258 (2000).

Sampson, D.A., Wang, M. and Matunis, M.J. The small ubiquitin-like modifier-1 (SUMO-1) consensus sequence mediates Ubc9 binding and is essential for SUMO-1 modification. *J Biol Chem* **276**, 21664-21669 (2001).

Sarge, K.D., Murphy, S.P. and Morimoto, R.I. Activation of heat shock gene transcription by heat shock factor 1 involves oligomerization, acquisition of DNA-binding activity, and nuclear localization and can occur in the absence of stress. *Mol Cell Biol* **13**, 1392-1407 (1993).

Sarge, K.D. and Park-Sarge, O.K. SUMO and its role in human diseases. *International review of cell and molecular biology* **288**, 167-183 (2011).

Schimmel, J., Eifler, K., Sigurethsson, J.O., Cuijpers, S.A., Hendriks, I.A., Verlaan-de Vries, M., Kelstrup, C.D., Francavilla, C., Medema, R.H., Olsen, J.V. and Vertegaal, A.C. Uncovering SUMOylation dynamics during cell-cycle progression reveals FoxM1 as a key mitotic SUMO target protein. *Mol Cell* **53**, 1053-1066 (2014).

Schimmel, J., Larsen, K.M., Matic, I., van Hagen, M., Cox, J., Mann, M., Andersen, J.S. and Vertegaal, A.C. The ubiquitin-proteasome system is a key component of the SUMO-2/3 cycle. *Mol Cell Proteomics* **7**, 2107-2122 (2008).

Schneider, C.A., Rasband, W.S. and Eliceiri, K.W. NIH Image to ImageJ: 25 years of image analysis. *Nat Methods* **9**, 671-675 (2012).

Schou, J., Kelstrup, C.D., Hayward, D.G., Olsen, J.V. and Nilsson, J. Comprehensive identification of SUMO2/3 targets and their dynamics during mitosis. *PLoS One* **9**, e100692 (2014).

Schulz, S., Chachami, G., Kozaczkiewicz, L., Winter, U., Stankovic-Valentin, N., Haas, P., Hofmann, K., Urlaub, H., Ovaa, H., Wittbrodt, J., Meulmeester, E. and Melchior, F. Ubiquitin-specific protease-like 1 (USPL1) is a SUMO isopeptidase with essential, non-catalytic functions. *EMBO Rep* **13**, 930-938 (2012).

Schwartz, Y.B. and Pirrotta, V. A new world of Polycombs: unexpected partnerships and emerging functions. *Nat Rev Genet* **14**, 853-864 (2013).

Seifert, A., Schofield, P., Barton, G.J. and Hay, R.T. Proteotoxic stress reprograms the chromatin landscape of SUMO modification. *Sci Signal* **8**, rs7 (2015).

Shalgi, R., Hurt, J.A., Krykbaeva, I., Taipale, M., Lindquist, S. and Burge, C.B. Widespread regulation of translation by elongation pausing in heat shock. *Mol Cell* **49**, 439-452 (2013).

Shalgi, R., Hurt, J.A., Lindquist, S. and Burge, C.B. Widespread inhibition of posttranscriptional splicing shapes the cellular transcriptome following heat shock. *Cell Rep* **7**, 1362-1370 (2014).

Shalizi, A., Gaudilliere, B., Yuan, Z., Stegmuller, J., Shirogane, T., Ge, Q., Tan, Y., Schulman, B., Harper, J.W. and Bonni, A. A calcium-regulated MEF2 sumoylation switch controls postsynaptic differentiation. *Science* **311**, 1012-1017 (2006).

Shen, L., Tatham, M.H., Dong, C., Zagorska, A., Naismith, J.H. and Hay, R.T. SUMO protease SENP1 induces isomerization of the scissile peptide bond. *Nat Struct Mol Biol* **13**, 1069-1077 (2006).

Shin, E.J., Shin, H.M., Nam, E., Kim, W.S., Kim, J.H., Oh, B.H. and Yun, Y. DeSUMOylating isopeptidase: a second class of SUMO protease. *EMBO Rep* **13**, 339-346 (2012).

Sievers, F., Wilm, A., Dineen, D., Gibson, T.J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Soding, J., Thompson, J.D. and Higgins, D.G. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol* **7**, 539 (2011).

Silen, J.L., Frank, D., Fujishige, A., Bone, R. and Agard, D.A. Analysis of prepro-alpha-lytic protease expression in Escherichia coli reveals that the pro region is required for activity. *J Bacteriol* **171**, 1320-1325 (1989).

Sloan, E., Tatham, M.H., Groslambert, M., Glass, M., Orr, A., Hay, R.T. and Everett, R.D. Analysis of the SUMO2 Proteome during HSV-1 Infection. *PLoS Pathog* **11**, e1005059 (2015).

Solis, E.J., Pandey, J.P., Zheng, X., Jin, D.X., Gupta, P.B., Airoldi, E.M., Pincus, D. and Denic, V. Defining the Essential Function of Yeast Hsf1 Reveals a Compact Transcriptional Program for Maintaining Eukaryotic Proteostasis. *Mol Cell* **63**, 60-71 (2016).

Song, J., Durrin, L.K., Wilkinson, T.A., Krontiris, T.G. and Chen, Y. Identification of a SUMO-binding motif that recognizes SUMO-modified proteins. *Proc Natl Acad Sci U S A* **101**, 14373-14378 (2004).

Song, J., Zhang, Z., Hu, W. and Chen, Y. Small ubiquitin-like modifier (SUMO) recognition of a SUMO binding motif: a reversal of the bound orientation. *J Biol Chem* **280**, 40122-40129 (2005).

Sorensen, S.B., Sorensen, T.L. and Breddam, K. Fragmentation of proteins by S. aureus strain V8 protease. Ammonium bicarbonate strongly inhibits the enzyme but does not improve the selectivity for glutamic acid. *FEBS Lett* **294**, 195-197 (1991).

Spriggs, K.A., Bushell, M. and Willis, A.E. Translational regulation of gene expression during conditions of cell stress. *Mol Cell* **40**, 228-237 (2010).

Sramko, M., Markus, J., Kabat, J., Wolff, L. and Bies, J. Stress-induced inactivation of the c-Myb transcription factor through conjugation of SUMO-2/3 proteins. *J Biol Chem* **281**, 40065-40075 (2006).

Steen, H. and Mann, M. The ABC's (and XYZ's) of peptide sequencing. *Nat Rev Mol Cell Biol* **5**, 699-711 (2004).

Steffan, J.S., Agrawal, N., Pallos, J., Rockabrand, E., Trotman, L.C., Slepko, N., Illes, K., Lukacsovich, T., Zhu, Y.Z., Cattaneo, E., Pandolfi, P.P., Thompson, L.M. and Marsh, J.L. SUMO modification of Huntingtin and Huntington's disease pathology. *Science* **304**, 100-104 (2004).

Steinacher, R. and Schar, P. Functionality of human thymine DNA glycosylase requires SUMO-regulated changes in protein conformation. *Curr Biol* **15**, 616-623 (2005).

Sternsdorf, T., Jensen, K. and Will, H. Evidence for covalent modification of the nuclear dot-associated proteins PML and Sp100 by PIC1/SUMO-1. *J Cell Biol* **139**, 1621-1634 (1997).

Streich, F.C., Jr. and Lima, C.D. Capturing a substrate in an activated RING E3/E2-SUMO complex. *Nature* **536**, 304-308 (2016).

Subramaniam, S., Sixt, K.M., Barrow, R. and Snyder, S.H. Rhes, a striatal specific protein, mediates mutant-huntingtin cytotoxicity. *Science* **324**, 1327-1330 (2009).

Swaney, D.L., Wenger, C.D. and Coon, J.J. Value of using multiple proteases for large-scale mass spectrometry-based proteomics. *J Proteome Res* **9**, 1323-1329 (2010).

Tammsalu, T., Matic, I., Jaffray, E.G., Ibrahim, A.F., Tatham, M.H. and Hay, R.T. Proteome-wide identification of SUMO2 modification sites. *Sci Signal* **7**, rs2 (2014).

Tammsalu, T., Matic, I., Jaffray, E.G., Ibrahim, A.F., Tatham, M.H. and Hay, R.T. Proteome-wide identification of SUMO modification sites by mass spectrometry. *Nat Protoc* **10**, 1374-1388 (2015).

Tanaka, K., Nishide, J., Okazaki, K., Kato, H., Niwa, O., Nakagawa, T., Matsuda, H., Kawamukai, M. and Murakami, Y. Characterization of a fission yeast SUMO-1 homologue, pmt3p, required for multiple nuclear events, including the control of telomere length and chromosome segregation. *Mol Cell Biol* **19**, 8660-8672 (1999).

Tatham, M.H., Geoffroy, M.C., Shen, L., Plechanovova, A., Hattersley, N., Jaffray, E.G., Palvimo, J.J. and Hay, R.T. RNF4 is a poly-SUMO-specific E3 ubiquitin ligase required for arsenic-induced PML degradation. *Nat Cell Biol* **10**, 538-546 (2008).

Tatham, M.H., Jaffray, E., Vaughan, O.A., Desterro, J.M., Botting, C.H., Naismith, J.H. and Hay, R.T. Polymeric chains of SUMO-2 and SUMO-3 are conjugated to protein substrates by SAE1/SAE2 and Ubc9. *J Biol Chem* **276**, 35368-35374 (2001).

Tatham, M.H., Matic, I., Mann, M. and Hay, R.T. Comparative proteomic analysis identifies a role for SUMO in protein quality control. *Sci Signal* **4**, rs4 (2011).

Tatham, M.H., Rodriguez, M.S., Xirodimas, D.P. and Hay, R.T. Detection of protein SUMOylation in vivo. *Nat Protoc* **4**, 1363-1371 (2009).

The Gene Ontology Consortium Gene Ontology Consortium: going forward. *Nucleic Acids Res* **43**, D1049-1056 (2015).

The Huntington's Disease Collaborative Research Group A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's disease chromosomes. . *Cell* **72**, 971-983 (1993).

The UniProt Consortium UniProt: a hub for protein information. *Nucleic Acids Res* **43**, D204-212 (2015).

Tirard, M., Hsiao, H.H., Nikolov, M., Urlaub, H., Melchior, F. and Brose, N. In vivo localization and identification of SUMOylated proteins in the brain of His6-HA-SUMO1 knock-in mice. *Proc Natl Acad Sci U S A* **109**, 21122-21127 (2012).

Troshin, P.V., Procter, J.B. and Barton, G.J. Java bioinformatics analysis web services for multiple sequence alignment--JABAWS:MSA. *Bioinformatics* **27**, 2001-2002 (2011).

Tsiatsiani, L. and Heck, A.J. Proteomics beyond trypsin. *Febs J* **282**, 2612-2626 (2015).

Udeshi, N.D., Mertins, P., Svinkina, T. and Carr, S.A. Large-scale identification of ubiquitination sites by mass spectrometry. *Nat Protoc* **8**, 1950-1960 (2013a).

Udeshi, N.D., Svinkina, T., Mertins, P., Kuhn, E., Mani, D.R., Qiao, J.W. and Carr, S.A. Refined preparation and use of anti-diglycine remnant (K-epsilon-GG) antibody enables routine quantification of 10,000s of ubiquitination sites in single proteomics experiments. *Mol Cell Proteomics* **12**, 825-831 (2013b).

Uhlen, M., Fagerberg, L., Hallstrom, B.M., Lindskog, C., Oksvold, P., Mardinoglu, A., Sivertsson, A., Kampf, C., Sjostedt, E., Asplund, A., Olsson, I., Edlund, K., Lundberg, E., Navani, S., Szigyarto, C.A., Odeberg, J., Djureinovic, D., Takanen, J.O., Hober, S., Alm, T., Edqvist, P.H., Berling, H., Tegel, H., Mulder, J., Rockberg, J., Nilsson, P., Schwenk, J.M., Hamsten, M., von Feilitzen, K., Forsberg, M., Persson, L., Johansson, F., Zwahlen, M., von Heijne, G., Nielsen, J. and Ponten, F. Proteomics. Tissue-based map of the human proteome. *Science* **347**, 1260419 (2015).

van Hagen, M., Overmeer, R.M., Abolvardi, S.S. and Vertegaal, A.C. RNF4 and VHL regulate the proteasomal degradation of SUMO-conjugated Hypoxia-Inducible Factor-2alpha. *Nucleic Acids Res* **38**, 1922-1931 (2010).

Van Nguyen, T., Angkasekwinai, P., Dou, H., Lin, F.M., Lu, L.S., Cheng, J., Chin, Y.E., Dong, C. and Yeh, E.T. SUMO-specific protease 1 is critical for early lymphoid development through regulation of STAT5 activation. *Mol Cell* **45**, 210-221 (2012).

Vassileva, M.T. and Matunis, M.J. SUMO modification of heterogeneous nuclear ribonucleoproteins. *Mol Cell Biol* **24**, 3623-3632 (2004).

Vertegaal, A.C., Andersen, J.S., Ogg, S.C., Hay, R.T., Mann, M. and Lamond, A.I. Distinct and overlapping sets of SUMO-1 and SUMO-2 target proteins revealed by quantitative proteomics. *Mol Cell Proteomics* **5**, 2298-2310 (2006).

Vertegaal, A.C., Ogg, S.C., Jaffray, E., Rodriguez, M.S., Hay, R.T., Andersen, J.S., Mann, M. and Lamond, A.I. A proteomic study of SUMO-2 target proteins. *J Biol Chem* **279**, 33791-33798 (2004).

Vethantham, V., Rao, N. and Manley, J.L. Sumoylation regulates multiple aspects of mammalian poly(A) polymerase function. *Genes Dev* **22**, 499-511 (2008).

Vijay-Kumar, S., Bugg, C.E. and Cook, W.J. Structure of ubiquitin refined at 1.8 A resolution. *J Mol Biol* **194**, 531-544 (1987).

Vizcaino, J.A., Csordas, A., del-Toro, N., Dianes, J.A., Griss, J., Lavidas, I., Mayer, G., Perez-Riverol, Y., Reisinger, F., Ternent, T., Xu, Q.W., Wang, R. and Hermjakob, H. 2016 update of the PRIDE database and its related tools. *Nucleic Acids Res* **44**, D447-456 (2016).

Wallace, E.W., Kear-Scott, J.L., Pilipenko, E.V., Schwartz, M.H., Laskowski, P.R., Rojek, A.E., Katanski, C.D., Riback, J.A., Dion, M.F., Franks, A.M., Airoldi, E.M., Pan, T., Budnik, B.A. and Drummond, D.A. Reversible, Specific, Active Aggregates of Endogenous Proteins Assemble upon Heat Stress. *Cell* **162**, 1286-1298 (2015).

Wang, L., Wansleeben, C., Zhao, S., Miao, P., Paschen, W. and Yang, W. SUMO2 is essential while SUMO3 is dispensable for mouse embryonic development. *EMBO Rep* **15**, 878-885 (2014).

Waterhouse, A.M., Procter, J.B., Martin, D.M., Clamp, M. and Barton, G.J. Jalview Version 2--a multiple sequence alignment editor and analysis workbench. *Bioinformatics* **25**, 1189-1191 (2009).

Wei, L. and Zhao, X. A new MCM modification cycle regulates DNA replication initiation. *Nat Struct Mol Biol* **23**, 209-216 (2016).

Wickham, H. (2009). ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag, New York, USA.

Wilhelm, M., Schlegl, J., Hahne, H., Moghaddas Gholami, A., Lieberenz, M., Savitski, M.M., Ziegler, E., Butzmann, L., Gessulat, S., Marx, H., Mathieson, T., Lemeer, S., Schnatbaum, K., Reimer, U., Wenschuh, H., Mollenhauer, M., Slotta-Huspenina, J., Boese, J.H., Bantscheff, M., Gerstmair, A., Faerber, F. and Kuster, B. Mass-spectrometry-based draft of the human proteome. *Nature* **509**, 582-587 (2014).

Wisniewski, J.R., Hein, M.Y., Cox, J. and Mann, M. A "proteomic ruler" for protein copy number and concentration estimation without spike-in standards. *Mol Cell Proteomics* **13**, 3497-3506 (2014).

Wisniewski, J.R., Zielinska, D.F. and Mann, M. Comparison of ultrafiltration units for proteomic and N-glycoproteomic analysis by the filter-aided sample preparation method. *Anal Biochem* **410**, 307-309 (2011).

Wisniewski, J.R., Zougman, A. and Mann, M. Combination of FASP and StageTip-based fractionation allows in-depth analysis of the hippocampal membrane proteome. *J Proteome Res* **8**, 5674-5678 (2009a).

Wisniewski, J.R., Zougman, A., Nagaraj, N. and Mann, M. Universal sample preparation method for proteome analysis. *Nat Methods* **6**, 359-362 (2009b).

Wohlschlegel, J.A., Johnson, E.S., Reed, S.I. and Yates, J.R., 3rd Improved identification of SUMO attachment sites using C-terminal SUMO mutants and tailored protease digestion strategies. *J Proteome Res* **5**, 761-770 (2006).

Wu, S.Y. and Chiang, C.M. Crosstalk between sumoylation and acetylation regulates p53-dependent chromatin transcription and DNA binding. *Embo J* **28**, 1246-1259 (2009).

Xu, G., Paige, J.S. and Jaffrey, S.R. Global analysis of lysine ubiquitination by ubiquitin remnant immunoaffinity profiling. *Nat Biotechnol* **28**, 868-873 (2010).

Yamashita, D., Yamaguchi, T., Shimizu, M., Nakata, N., Hirose, F. and Osumi, T. The transactivating function of peroxisome proliferator-activated receptor gamma is negatively regulated by SUMO conjugation in the amino-terminal domain. *Genes Cells* **9**, 1017-1029 (2004).

Yeh, E.T., Gong, L. and Kamitani, T. Ubiquitin-like proteins: new wines in new bottles. *Gene* **248**, 1-14 (2000).

Yun, S.M., Cho, S.J., Song, J.C., Song, S.Y., Jo, S.A., Jo, C., Yoon, K., Tanzi, R.E., Choi, E.J. and Koh, Y.H. SUMO1 modulates Abeta generation via BACE1 accumulation. *Neurobiol Aging* **34**, 650-662 (2013).

Zhang, F.P., Mikkonen, L., Toppari, J., Palvimo, J.J., Thesleff, I. and Janne, O.A. Sumo-1 function is dispensable in normal mouse development. *Mol Cell Biol* **28**, 5381-5390 (2008).

Zhang, Y.Q. and Sarge, K.D. Sumoylation of amyloid precursor protein negatively regulates Abeta aggregate levels. *Biochem Biophys Res Commun* **374**, 673-678 (2008).

Zhao, Y., Kwon, S.W., Anselmo, A., Kaur, K. and White, M.A. Broad spectrum identification of cellular small ubiquitin-related modifier (SUMO) substrate proteins. *J Biol Chem* **279**, 20999-21002 (2004).

Zhou, J., Wan, J., Gao, X., Zhang, X., Jaffrey, S.R. and Qian, S.B. Dynamic m(6)A mRNA methylation directs translational control of heat shock response. *Nature* **526**, 591-594 (2015).

Zoghbi, H.Y. and Orr, H.T. Glutamine repeats and neurodegeneration. *Annu Rev Neurosci* **23**, 217-247 (2000).

6.

Appendix

6.1. Supplementary files and tables

The mass spectrometry proteomics data associated with publications provided in Appendix 6.2 (Tammsalu et al., 2014; Kohler et al., 2015; Tammsalu et al., 2015; Pelisch et al., 2016) have been deposited to the ProteomeXchange Consortium (http://proteomecentral.proteomexchange.org) via the PRIDE partner repository (Vizcaino et al., 2016) with data set identifiers PXD001281, PXD002972 and PXD005202.

Unpublished raw MS and MaxQuant output files, and supplementary materials are available upon request.

6.2. Publications

The following research articles have been published during the course of this work:

Tammsalu, T.*, Matic, I.*, Jaffray, E.G., Ibrahim, A.F., Tatham, M.H. and Hay, R.T. Proteome-wide identification of SUMO2 modification sites. *Sci Signal* **7**, rs2 (2014).

Tammsalu, T.*, Matic, I.*, Jaffray, E.G., Ibrahim, A.F., Tatham, M.H. and Hay, R.T. Proteome-wide identification of SUMO modification sites by mass spectrometry. *Nat Protoc* **10**, 1374-1388 (2015).

Kohler, J.B.*, **Tammsalu, T.***, Jorgensen, M.M., Steen, N., Hay, R.T. and Thon, G. Targeting of SUMO substrates to a Cdc48-Ufd1-Npl4 segregase and STUbL pathway in fission yeast. *Nat Commun* **6**, 8827 (2015).

Pelisch, F., **Tammsalu, T.**, Wang, B., Jaffray, E.G., Gartner, A. and Hay, R.T. A SUMO-Dependent Protein Network Regulates Chromosome Congression during Oocyte Meiosis. *Mol Cell* (2016).

*Contributed equally to the work.